UNIVERSITÉS FRANCOPHONES

André Clas
Université de Montréal

Hayssam Safar
Université de Mons-Hainaut

L'ENVIRONNEMENT TRADUCTIONNEL

La station de travail du traducteur de l'an 2001

Actes du Colloque de Mons 1991

AUPELF

actualité scientifique

Presses de l'Université du Québec

L'ENVIRONNEMENT TRADUCTIONNEL

La station de travail du traducteur de l'an 2001

Données de catalogage avant publication (Canada)

Réseau thématique de recherche « Lexicologie, terminologie, traduction ». Journées scientifiques (2e : 1991 : Mons, Belgique)

L'environnement traductionnel : la station de travail du traducteur de l'an 2001 : Journées scientifiques du Réseau thématique de recherche « Lexicologie, terminologie, traduction », Mons, 25-27 avril 1991

(Universités francophones. Actualité scientifique)
Publ. en collab. avec : UREF et AUPELF.
Comprend un index.

ISBN 2-7605-0705-X

1. Traduction automatique – Congrès. 2. Traduction – Logiciels – Congrès. 3. Dictionnaires électroniques – Congrès. 4. Terminologie – Informatique – Congrès. 5. Traducteurs – Formation – Congrès. 6. Traduction, Services de – Congrès. I. Clas, André, 1933- . II. Safar, Hayssam. III. UREF. IV. Association des universités partiellement ou entièrement de langue française. V. Titre. VI. Collection.

P308.J68 1992 418'.02'0285 C92-096607-1

ISBN 2-7605-0705-X

Tous droits de reproduction, de traduction et d'adaptation réservés © 1992
Presses de l'Université du Québec

Dépôt légal – 2e trimestre 1992
Bibliothèque nationale du Québec
Bibliothèque nationale du Canada
Imprimé au Canada

L'ENVIRONNEMENT TRADUCTIONNEL

La station de travail du traducteur de l'an 2001

Journées scientifiques du Réseau thématique de recherche
« Lexicologie, terminologie, traduction »
Mons 25-27 avril 1991

Sous la direction de

André CLAS, Université de Montréal
Hayssam SAFAR, Université de Mons-Hainaut

1992

Presses de l'Université du Québec
Case Postale 250
Sillery (Québec) Canada
G1T 2R1

AUPELF • UREF
B.P. 400, succ. Côte-des-Neiges
Montréal (Québec) Canada
H3S 2S5

Les collections « Universités francophones » de l'UREF : Un instrument vital pour l'évolution de l'espace scientifique francophone

L'Université des réseaux d'expression française (UREF) créée au sein de l'Association des universités partiellement ou entièrement de langue française (AUPELF) lors du Sommet des Chefs d'État et de Gouvernement des pays ayant en commun l'usage du français, à Québec, en 1987, a développé des collections de manuels et ouvrages scientifiques de haut niveau. Regroupées sous le nom « Universités francophones », ces collections sont les nouveaux outils de renforcement de l'espace scientifique en français, enjeu primordial pour l'avenir de la francophonie.

Dans le même temps, l'UREF a mis en place des réseaux thématiques de recherche, rassemblant de façon multilatérale plusieurs milliers d'enseignants et de chercheurs de toute la francophonie. À ce jour, 15 réseaux de recherche favorisent la production et l'échange d'information scientifique et technique dans les domaines, notamment, de la médecine, du droit et des sciences de la gestion, de l'environnement, de l'agronomie, de la télédétection, des sciences humaines.

La série « Actualité scientifique » dans laquelle s'inscrit le présent ouvrage, au sein d'« Universités francophones », est constituée des actes des colloques et journées scientifiques des réseaux de l'UREF. Elle accueille ainsi, en français, le bilan des travaux de recherche dans les domaines d'activités scientifiques de ces réseaux.

Une série de manuels universitaires ainsi qu'une série « Sciences en marche » constituée de monographies de recherche, complètent le dispositif éditorial de l'UREF qui comprend, par ailleurs, des revues scientifiques (les *Cahiers d'études et de recherches francophones* – 2 titres : *Santé, Agricultures; Science et changements planétaires, Sécheresse*).

Nous comptons bien que ces instruments, qui sont mis à la disposition de nos partenaires des pays du Sud à un prix préférentiel, contribueront à la consolidation d'une francophonie scientifique soucieuse d'excellence et de rayonnement international.

Professeur Michel Guillou
Recteur de l'UREF

Sommaire

Liste des auteurs XI
Membres du comité du réseau « LTT » XIII
Allocution. *Marcel Voisin* XV
Allocution. *André Clas* XIX

Partie I. **Traduction automatique et traduction assistée**

1. **Quelle automatisation de la traduction peut-on souhaiter et réaliser sur les stations de travail individuelles?** *C. Boitet (Grenoble)* 3
2. **L'environnement sémantique, aide au transfert lexical en TA. Problèmes posés par le calcul des traductions françaises de quelques verbes anglais** *A. Attali, A. Euvrard (Nancy)* 21
3. **Problèmes de désambiguïsation interactive en TAO personnelle** *H. Blanchon (Grenoble)* 31
4. **Dictionnaires électroniques dans un environnement traductionnel avancé** *U. Heid (Stuttgart)* 49
5. **Le module lexical dans un système de traduction automatique** *L. Degand, G. Everaert, J. Lerot (Louvain-la-Neuve)* 51
6. **Pour une approche interactive au problème de la traduction automatique** *E. Wehrli (Genève)* 59
7. **La traduction automatique des bulletins d'avalanches de la Suisse** *P. Bouillon, K. Boesefeldt (Genève)* 69
8. **Le projet EUROTRA, projet européen de recherche sur la traduction automatique** *C. Jadelot, D. Boussel (Nancy)* 79
9. **ODILE 2, un outil pour traducteurs occasionnels sur Macintosh** *M. Lafourcade (Grenoble)* 95
10. **Intégration des connaissances en génie linguistique : niveaux, dimensions, objets et contraintes** *J.-Y. Morin (Montréal)* 109

11. Poste de travail de traduction par l'informatisation d'un dictionnaire français-arabe *S. Taleb (Rabat)* 135
12. Traduction et terminologie : expérience et perspectives industrielles *D. Bachut, R. Gerber (Grenoble)* 139

Partie II. **La station de travail du traducteur**

13. Pour le traducteur : un poste de travail à trois niveaux d'assistance *A. Melby (Provo)* 151
14. Le poste de travail de traducteur à l'ONU — Horizon 2001 *M. Gainet (New York)* 155
15. La technologie au service du traducteur *K. Tremblay (Ottawa)* 161
16. PTT-2 : la deuxième version du poste de travail du traducteur du CCRIT *E. Macklovitch (Laval)* 171
17. Trois scénarios possibles pour la station de travail du traducteur de l'an 2001 *P. Søren Kjærsgaard (Odense)* 183
18. Un projet de dictionnaire intelligent *J.-M. Maes (Grand)* 189
19. EURODICAUTOM, ou la terminologie en l'an 2001 *A. Reichling (Luxembourg)* 201
20. La modernisation de la traduction à la Commission et ses moyens *K. Van Der Horst (Luxembourg)* 207

Partie III. **Traduction, Terminologie, Informatique**

21. Aspects de l'informatisation des activités terminologiques et traductionnelles *G. Otman (Paris)* 213
22. Comparaison de logiciels utilisables en terminologie *É. Blanchon (Paris)* 223
23. Stratégies de professionnalisation de la formation des traducteurs *D. Gouadec (Rennes)* 235
24. Quelques logiciels utiles aux traducteurs ou aux terminologues. Les hypertextes et l'analyse des contextes *C. Doutrelepont (Ottawa)* 249
25. Forme d'un dictionnaire électronique *G. Gross (Villetaneuse)* 255
26. À la recherche d'écosystèmes terminologiques *D. Blampain, P. Petrussa, M. Van Campenhoudt (Bruxelles)* 273
27. Des fichiers terminologiques aux bases de connaissances *B. De Bessé (Genève)* 283
28. Importance de la traduction dans l'aménagement linguistique de la République centrafricaine *M. Diki-Kidiri (Paris)* 301

29. Jargonautes et terminophages ou le traducteur, spécialiste de la langue, au service du consommateur, handicapé linguistique *L. Y. Chaballe, J. Klein (Mons)* 305
30. La station de travail du traducteur professionnel et ses implications pédagogiques *J. Klein, L. Y. Chaballe (Mons)* 315
31. En dessous de quel seuil de qualité... *J.-M. Waaub (Mons)* 321
32. L'ordinateur est une vraie machine à écrire *J.-R. Ladmiral (Paris)* 329
33. La formation du traducteur en l'an 2001 *K. Dejean Le Féal (Paris)* 341
34. Former des formateurs de traducteurs pour le XXI^e siècle. Pour un stage intégré et francophone de traduction *J.-C. Gémar (Montréal)* 349

Partie IV. **Tables rondes**

35. Première table ronde : Recherche et formation 359
36. Deuxième table ronde : Problèmes et besoins pour l'Afrique 361
37. Troisième table ronde : Perspectives et stratégies de développement de la traduction dans le monde arabe 363

Index 371

Liste des auteurs

Amid Abdallah UNESCO, Paris, France.

Attali Arlette, Euvrard Annette CELTA, CRAL, CNRS / EUROTRA F-Nancy, Université de Nancy II, BP 3397, 54015 Nancy Cedex, France.

Bachut Daniel, SITE, 12, rue de Reims, 94701 Maisons-Alfort Cedex, France.

Blampain Daniel, Petrussa Philippe, Van Campenhoudt Marc Institut supérieur des traducteurs et interprètes, Communauté française, Bruxelles, Belgique.

Blanchon Élisabeth Centre de Terminologie et de Néologie, CNRS, INaLF, 27, rue Damesme, 75013 Paris, France.

Blanchon Hervé GETA, Institut IMAG (UJF & CNRS), BP 53X, 38041 Grenoble Cedex, France.

Boitet Christian GETA, Institut IMAG (UJF & CNRS), BP 53X, 38041 Grenoble Cedex, France.

Bouillon Pierrette, Boesefeldt Katharina ISSCO, 54, route des Acacias, 1207 Genève, Suisse.

Chaballe Louis Yvon Chaballe Traduction & Communication, Mons, Belgique.

Clas André Coordonnateur du réseau, professeur, directeur du GRESLET, Université de Montréal, CP 6128, Succursale A, Montréal, Québec, Canada.

de Bessé Bruno École de traduction et d'interprétation, Université de Genève, CH-1205, Genève, Suisse.

Degand Liesbeth, Everaert Guy, Lerot Jacques Projet GENESE, Institut de linguistique, Place Blaise Pascal, 1, B-1348 Louvain-la-Neuve, Belgique.

Dejean Le Féal Karla École supérieure d'interprètes et de traducteurs, Université de Paris 3, Paris, France.

Diki-Kidiri Marcel CNRS (LACITO), Paris, France.

Doutrelepont Charles Département de français, Université Carleton, Ottawa, Ontario, K1S 5B6, Canada.

Gainet Michel Président du Groupe de travail des innovations technologiques, Division de traduction, Organisation des Nations Unies, Bureau S-1294, New York, NY 10017, États-Unis.

Gémar Jean-Claude Département de linguistique et philologie, Université de Montréal, CP 6128, Succursale A, Montréal, Québec, Canada.

Gerber René B'VITAL, 35, rue Joseph Chanrion, 38000 Grenoble, France.

Gouadec Daniel Formation des Traducteurs, UFR de Langues Appliquées, Université de Rennes 2, 35043 Rennes Cedex, France.

Gross Gaston Laboratoire de linguistique informatique, Université Paris 13, Villetaneuse, France.

Hardane Jarjoura École de traducteurs et d'interprètes, Université Saint-Joseph, Beyrouth, Liban.

Heid Ulrich Université de Stuttgart, Stuttgart, Allemagne.

Jadelot Christiane, Boussel Dominique CELTA, CNRS / EUROTRA F-Nancy, Université de Nancy 2, BP 3397, 54015 Nancy Cedex, France.

Kjærsgaard Poul Søren Université d'Odense, Campusvej 55, DK-5230 Odense M., Danemark.

Klein Jean École d'Interprètes Internationaux de l'Université de Mons-Hainaut, Mons, Belgique.

Ladmiral Jean-René Université de Parie-X-Nanterre, Centre d'études et de recherches en traduction, Institut supérieur d'interprétation et de traduction, Paris, France.

Lafourcade Mathieu GETA, Institut IMAG (UJF & CNRS), Grenoble, France.

Macklovitch Elliott Groupe de traduction assistée, Centre canadien de recherche sur l'informatisation du travail, 1575, boul. Chomedey, Laval, Québec, H7V 2X2, Canada.

Maes Jean-Marie Provinciaal Instituut Voor Hoger Onderwijs (PIHO), Henleykaai 84, B-9000 Gand, Belgique.

Melby Alan Department of linguistics, 2129 JKHB, Brigham Young University at Provo, Provo, Utah 84602, États-Unis.

Morin Jean-Yves Université de Montréal, Département de linguistique et philologie, CP 6128, Succursale A, Montréal, Québec, H3C 3J7, Canada.

Otman Gabriel Centre de Terminologie et de Néologie, CNRS, INaLF, 27, rue Damesme, 75013 Paris, France.

Reichling Alain Commission des Communautés européennes, Service de traduction, L-2920 Luxembourg.

Taleb Saadia Responsable du département Bases de Données, Institut d'Études et de Recherches pour l'Arabisation (IERA), Université Mohamed V, BP 6216 Instituts, Rabat, Maroc.

Thoiron Philippe Directeur du CRTT, Université Lumière Lyon 2, Lyon, France.

Tremblay Klaire Chargée de projets TAO, Secrétariat d'État du Canada, Langues officielles et Traduction, Ottawa, Ontario, K1A 0M5, Canada.

Van Der Horst Kees Commission des Communautés Européennes, Service de traduction, Unité modernisation des méthodes de travail, L-2920 Luxembourg.

Verheve Daniel Faculté des sciences économiques et sociales, Université de Mons-Hainaut, Mons, Belgique.

Voisin Marcel Professeur, directeur de l'École d'Interprètes Internationaux de l'Université de Mons-Hainaut, Mons, Belgique.

Waaub Jean-Marie Université de Mons-Hainaut, Mons, Belgique.

Wehrli Eric Département de linguistique générale et de linguistique française, Université de Genève, 1211 Genève 4, Suisse.

Membres du comité du réseau « LTT »

Chad M. Professeur, doyen de la faculté des lettres, Université Sidi Mohamed Ben Abdallah, Fès, Maroc.

Clas, A. Coordonnateur du réseau, professeur, directeur du GRESLET, Université de Montréal, Montréal, Canada.

Ouoba, B. Professeur, Université de Ouagadougou, Burkina Faso.

Thoiron Ph. Professeur, directeur du CRTT, Université Lumière-Lyon 2, Lyon, France.

Allocution de M. Marcel VOISIN

Marcel VOISIN

Directeur de l'École d'Interprètes Internationaux, Université Mons-Hainaut, Mons, Belgique

L'origine de la traduction se perd dans la nuit des temps au fil duquel sa fonction n'a cessé d'être diversement ressentie.

La célèbre pierre de Rosette datée de 196 avant J.-C. ou la traduction latine de l'Ancien Testament à la fin du IV^e^ siècle par saint Jérôme — devenu le saint patron des traducteurs — sont, parmi beaucoup d'autres, des repères pour notre histoire.

Pendant longtemps, on s'est improvisé traducteur, par hasard, nécessité ou vocation, sans préparation autre que circonstancielle et sans entraînement que celui du terrain, le plus souvent occasionnel sans plus.

Mais cela n'empêchait pas d'en reconnaître le mérite ou l'utilité.

Ainsi Mme de Staël en 1816 écrit : « Il n'y a pas de plus éminent service à rendre à la littérature que de transporter d'une langue à l'autre les chefs-d'œuvre de l'esprit humain. » Et devançant notre modernité, elle ajoute : « D'ailleurs la circulation des idées est, de tous les genres de commerce, celui dont les avantages sont les plus certains[1]. »

En effet, aujourd'hui que la Terre devient un « village planétaire », le rôle vital des communications et des échanges est reconnu par tous. D'ores et déjà, la traduction tisse en ce domaine notre présent et notre avenir.

Dès lors, sortant de sa marginalité et de son anonymat, elle évolue avec toutes les disciplines qui bâtissent le futur et l'essor technologique lui impose sa loi. La voici écartelée entre son passé artisanal et un avenir industriel.

1. « De l'esprit des traductions », *Œuvres complètes*, Paris, 1821, t. XVII, p. 396.

Nous vivons cette charnière qui pose à l'enseignement des problèmes aigus. Comment concilier une tradition éprouvée avec une modernisation nécessaire? Comment une institution, par nature conservatrice, aux moyens limités, pourra-t-elle suivre l'accélération, quelque peu démentielle, de l'innovation technique? Comment « raison garder » entre la culture et la machine?

C'est un tel défi que nous tentons ce jour de relever, à tout le moins d'éclairer, avec l'aide de l'AUPELF et de l'UREF, grâce à la contribution savante ou à l'expérience d'orateurs venus d'une quinzaine de pays, s'adressant, du Nord au Sud et de l'Est à l'Ouest, à près d'une trentaine de nations ici représentées et, par la publication qui suivra, à la francophonie mondiale.

Des chercheurs, des spécialistes, des enseignants et des professionnels vont donc dialoguer pour le meilleur profit de tous. Ils vont esquisser la prospective du métier, la futurologie de la discipline, partant les nouvelles modalités de la formation. Ce dont je les remercie par avance.

En quatre années d'études, dès la sortie du secondaire pour la plupart — encore un défi! —, nous devrions former, non pas des spécialistes achevés, mais des généralistes qui soient à la fois des artisans et des industriels en puissance.

Que le traducteur se libère des servitudes du métier, des tâches répétitives et des recettes à l'instar de nombre de professions modernes, bravo! Mais qu'il ne devienne pas un O.S. de la traduction...

L'acte particulièrement humanisant de la communication interculturelle doit garder sa dimension esthétique, sa créativité valorisante, sa plénitude heureuse.

Sur ce point, l'exercice, particulièrement subtil et complet, de la traduction littéraire me paraît un entraînement essentiel, même si son usage demeure forcément limité. C'est sa difficulté même qui est révélatrice et précieuse. Et comme l'écrivait plaisamment un anonyme en 1836 : « En vérité, quand on passe en revue toutes les difficultés de l'art de traduire, on conçoit difficilement que l'on ose tenter l'entreprise[2]. » S'y jeter par inconscience serait catastrophique.

C'est au contraire une haute conscience éclairée, mue par la volonté d'apprendre sans cesse, dynamisée par le défi qu'il faudrait développer. Une haute pédagogie de l'obstacle en quelque sorte...

Dans ce domaine comme dans tant d'autres, nous vivrons la coexistence des produits industriels, standardisés, fiables, avec ceux de l'artisanat aux charmes surannés mais indispensables à notre art de vivre, à notre bonheur.

D'un côté rapidité, efficacité, rendement; de l'autre la poétique[3], cette dimension qui est le sel de notre être, donne sens à notre vie et rend passionnant le métier.

2. « Des traductions », *Bibliothèque Universelle de Genève*, juin 1836, t. III, p. 245.

3. La théorie romantique du génie créateur porte ombrage à la traduction qui pourtant, quand elle est réussie, est une véritable recréation où intervient la poétique du traducteur. Le XVIII[e] siècle l'avait bien aperçu.

Ainsi d'Alembert n'hésite pas à classer les traducteurs « immédiatement après les écrivains créateurs ». Son contemporain Beauzée écrit en 1765 : « Rien n'est plus difficile ni plus rare que de garder un juste milieu entre la licence du commentaire et la servitude de la lettre. » De même, Maximilien-Henri de Saint Simon dans son *Essai de traduction littérale et énergique* (sic!) de 1771 rapproche aussi le génie du traducteur de celui de l'auteur.

Nous souhaitons donc que ces journées scientifiques éclairent la double formation du traducteur de façon optimale et lui facilitent l'entrée de plain-pied dans le XXI[e] siècle. Nous avons besoin de synergies conquérantes et de dialectiques harmonieuses.

Merci à tous et à chacun de contribuer au mieux durant ces trois journées privilégiées qui vont nous faire vivre au cœur de la francophonie mondiale.

Allocution de M. André CLAS

André CLAS

Coordonnateur du réseau Lexicologie, Terminologie, Traduction, UREF, Université de Montréal, Montréal, Canada

Monsieur le Recteur,
Monsieur le Bourgmestre,
Messieurs les représentants des Gouvernements,
Mesdames et Messieurs,
Chers Collègues,

L'importance de la traduction n'est plus à souligner, et d'autant moins en Belgique, ce pays qui héberge une entité administrative où la traduction joue un rôle capital. Il faut, nous en sommes persuadé, accorder foi à la déclaration, il y a déjà quelque temps, d'un premier secrétaire des Nations Unies, qui affirmait que le sort du monde dépend autant des hommes politiques que des traducteurs. En fait, les hommes politiques ne peuvent se passer des services du traducteur pour présenter leurs idées, leurs propositions, pour comprendre les autres et se faire comprendre dans les réunions internationales. Nous vivons de plus en plus dans un espace interpolitique, intercommercial, interindustriel, interlinguistique et interculturel. Il faut non seulement communiquer avec l'Autre, il faut aussi savoir recevoir de l'Autre ce qui est important et utile pour soi. C'est un fait que l'on traduit depuis toujours et que, de nos jours, on traduit de plus en plus, et même on retraduit ce qui avait été traduit auparavant.

Si l'on traduit depuis fort longtemps, c'est qu'il y avait toujours volonté ou nécessité de comprendre, de savoir, de comparer, d'imiter, de s'inspirer, de faire mieux, de parfaire. La diversité des langues et des conditions de vie nourrit la diversité des civilisations. La comparaison s'impose et crée par répercussion une réflexion, une

découverte, une pensée neuve. Une communication interculturelle s'établit et permet à l'homme d'ajouter à son être un trait qui le définit mieux. Il y a là une fonction proligère de prime importance.

Si l'on doit, à cause de la rapidité des communications entre tous les pays, du volume des informations et surtout de la modification ultrarapide des données techniques et scientifiques, traduire de plus en plus et de plus en plus rapidement, il faut accorder une attention particulière à l'aide que la technique peut apporter au traducteur pour l'exécution de sa tâche dans des conditions optimales. Le traducteur, on l'a dit mais il convient de le répéter, joue un très grand rôle dans la diffusion des valeurs culturelles, dans la prise en charge des découvertes scientifiques et des réalisations techniques, bref dans le développement économique, social de l'être humain. Le traducteur est un maillon dans la chaîne des artisans du progrès, il est un diffuseur d'idées et de connaissances, et, ne l'oublions pas, un important responsable de la qualité de la langue.

Cette rencontre revêt donc une très grande importance pour nous tous, et c'est pour cela que le réseau LTT de l'UREF a inscrit cette manifestation à son programme. Les questions débattues ici sont importantes et incontournables. Il fallait une telle rencontre pour aller de l'avant, pour mieux être présent dans le monde de demain.

Permettez-moi tout d'abord de vous exprimer à tous mes remerciements pour être venus si nombreux. J'en suis très heureux et flatté; cela montre que nous avions besoin de nous voir, de dialoguer, de réfléchir ensemble pour savoir ce que l'Autre fait, ce que nous avons à faire, ce que nous pouvons faire ensemble.

Nous avons intitulé nos *Deuxièmes Journées scientifiques de notre réseau LTT* : « L'environnement traductionnel. La station de travail du traducteur de l'an 2001 » parce que nous voulions rappeler que la traduction est une profession de plus en plus indispensable de nos jours et qui dicte de façon primordiale la rencontre et la compréhension des hommes, le cheminement vers le progrès, le nouveau, le renouveau, le développement, la connaissance, le savoir, la prise de conscience de soi et des autres. Le monde est polyglotte, on le sait, mais cette particularité fondamentale nous oblige, puisque nous vivons dans un univers presque totalement interconnecté où ce qui se passe ici ne peut être ignoré ailleurs et ici, à traduire et à traduire de plus en plus et de plus en plus vite. La demande pour des traductions rapides, bien faites évidemment, et au prix de revient le plus bas possible, pose bien entendu de redoutables problèmes dont la solution va indéniablement vers un poste de travail de traducteur très sophistiqué, c'est-à-dire vers cette station de travail très performante que l'on peut imaginer très facilement, suivant divers scénarios. Peut-être avec une priorité accordée à la présence du traducteur, c'est lui l'artisan de la machine, ou peut-être seulement en lui accordant un rôle de surveillant, c'est la machine qui est l'artisan. On peut aussi entrevoir un cheminement qui va de la traduction banalement informatisée — l'utilisation d'un système quelconque de traitement de texte additionné à des logiciels de correction d'orthographe plus ou moins perfectionnés, à des conjugueurs, à des dictionnaires automatiques, à des systèmes d'aide à la rédaction — à l'ingérence de plus en plus importante des traitements automatiques et automatisés, c'est-à-dire qui s'appuie, à un degré plus ou moins marqué, sur l'intervention humaine directe.

On sait que la traduction automatique n'est pas idée nouvelle, pas plus que la traduction assistée, d'ailleurs. On peut, en schématisant beaucoup, diviser l'histoire de la traduction automatique en diverses périodes : en une première époque, vers les années cinquante — rappelons que la première construction de la « machine à traduire » remonte à 1946 —, marquée par un débordement des recherches, surtout entre 1957 et 1965, en une deuxième période qui est surtout celle des déceptions, des remises en cause, c'est la période des effets du rapport ALPAC, et en une troisième période, celle des années qui commencent en soixante-dix, et qui ordonnent les recherches de façon peut-être plus réaliste, l'ère du simple transcodage est terminée, la traduction est plus complexe qu'on ne le croyait. Les diverses recherches ont apporté des progrès sensibles dans divers domaines et ont aussi donné naissance à de multiples systèmes, plus ou moins anciens, plus ou moins performants. Nous connaissons tous des noms comme SYSTRAN, GETA, TAUM-METEO, METAL, SPANAM, SMART, LOGOS, SUSY, DLT, EUROTRA, pour n'en citer que quelques-uns. Parallèlement, il y a eu, vers les années soixante, des créations de ce qu'on a convenu d'appeler des « banques de terminologie », sur grands systèmes informatiques d'abord, puis, avec les progrès techniques, sur « systèmes personnalisés ». Les « aides » à la traduction sont de plus en plus nombreuses, les dictionnaires sont de plus en plus électroniques, tout est de plus en plus informatisé. Nous devons tous tirer avantage de ces réalisations, peut-être aussi dire ce que nous aimerions avoir, vers quoi nous devrions tendre. C'est tout cela que nous voulons explorer avec vous pour savoir où nous en sommes, pour savoir aussi, on ne le sait pas toujours, qui fait quoi et où, et aussi vers quoi nous irons, vers quoi nous voulons aller.

La sagesse des nations nous apprend que si nous avons des invités de bien s'occuper d'eux, de bien les recevoir, de bien les traiter, mais de les emmener le troisième jour aux champs et de donner à chacun une houe. Mesdames et Messieurs, nous sommes arrivés au troisième jour et je déclare ouvertes les *Deuxièmes Journées scientifiques du réseau LTT*.

PARTIE I

Traduction automatique et traduction assistée

Présidents : Marcel Voisin
Christian Delcourt
Roger Goffin

1

Quelle automatisation de la traduction peut-on souhaiter et réaliser sur les stations de travail individuelles?

Christian BOITET

GETA, Institut IMAG (UJF et CNRS), Grenoble, France

Résumé

Une brève revue de la situation actuelle permet de dégager une classification des types de traduction (re-création, localisation, traduction-diffusion, traduction-dépistage) et de déterminer dans quelle mesure on peut automatiser la « fonction traduisante » proprement dite, c'est-à-dire la production automatique de traductions « brutes » ou « grossières » (destinées à la diffusion après révision ou au dépistage). Comme les possibilités et les limites de la TAO « classique » (TA) sont souvent méconnues, nous indiquerons au passage quelques règles de « bon usage » de la TA.

En ce qui concerne les « stations de travail » associées, l'accès à des traductions grossières (dépistage) doit pouvoir être réalisé sur des matériels très variés, allant du terminal Minitel à un poste dédié. Par contre, la révision de traductions brutes nous semble demander le même type d'environnement que la traduction professionnelle industrielle (en groupe).

Quand on ne peut raisonnablement proposer une technique de TA, on peut souvent fournir des aides au traducteur humain. Dans un cadre professionnel, il faut prévoir un environnement pour le travail en petite équipe (*groupware*), l'accès aux traductions passées, l'intégration dans un système de documentation informatisée, et de puissantes bases de données lexicales et terminologiques. Les stations de travail et les logiciels utilisés peuvent être de haut de gamme, et/ou dédiés.

S'il s'agit de traduction occasionnelle, il faut des environnements moins ambitieux, mais utilisables en conjonction avec une grande variété d'outils du commerce (texteurs, documenteurs, tableurs, grapheurs, SGBD...). Du point de vue informatique, la conception et la réalisation de ce type d'outils est particulièrement intéressante, à cause des exigences de généricité, de portabilité et de simplicité. Bien sûr, il faut viser des matériels accessibles au grand public.

À côté des stations pour le veilleur, pour le réviseur-traducteur professionnel, et pour le traducteur occasionnel, on peut enfin imaginer (mais sera-ce pour l'an 2001?) des stations pour rédacteurs unilingues ne pouvant utiliser les services d'un traducteur, désirant cependant être traduits dans une ou plusieurs langues, et acceptant pour cela de négocier et de clarifier leurs textes par dialogue avec le système.

Mots clés

Traduction assistée par ordinateur (TAO), Traduction automatique (TA), Traduction humaine assistée par machine (THAM), traduction-dépistage, traduction-diffusion, traduction professionnelle, traduction occasionnelle, stations de travail individuelles pour la traduction.

Introduction

On emploie le terme de « traduction » aussi bien pour la poésie que pour les romans, la publicité, les ouvrages scientifiques, les rapports et manuels techniques, les nomenclatures de pièces détachées, alors qu'il conviendrait, au moins, de distinguer :

- la « re-création », par exemple la traduction d'Edgar Allan Poe par Baudelaire, qui vise avant tout à transmettre l'aspect subjectif, fût-ce au prix d'une légère transformation du contenu;
- la « localisation », largement pratiquée pour les manuels de micro-ordinateurs, qui vise à adapter un contenu à un environnement culturel particulier;
- la « traduction-diffusion » [1.14], en particulier la traduction de documentations techniques dont le contenu doit être strictement rendu, sans ajout ni omission, même si le style « sent la traduction »; et
- la « traduction rapide » enfin, dans laquelle nous rangerons la « traduction-dépistage » de textes écrits et l'interprétation simultanée.

Quelle automatisation de la traduction peut-on souhaiter et réaliser sur des stations de travail individuelles dans les dix prochaines années? Après avoir examiné les types d'automatisation envisageables, et les situations traductionnelles, qui ne se limitent pas à l'image d'Épinal d'une « photocopieuse-traductrice », nous conclurons qu'il vaut mieux en général ne pas parler de « stations de travail », mais plutôt d'environnements ou d'outils, qui ne pourront être spécifiques que dans un cas, celui de la traduction professionnelle en groupe, avec ou sans traduction automatique.

La TAO actuelle

Il n'est absolument pas envisageable pour l'instant d'automatiser la traduction-recréation ni la traduction-localisation plus que par la mise à disposition d'outils d'aide au traducteur humain. Par contre, la « fonction traduisante » est automatisable pour la traduction-diffusion et la traduction-dépistage.

Le terme « TAO » (Traduction assistée par ordinateur) recouvre aujourd'hui l'ensemble des techniques d'automatisation de la traduction. L'ancien terme « TA » (Traduction automatique) est réservé aux techniques d'automatisation de la fonction traduisante, qu'il y ait ou non préédition, postédition ou interaction, tandis que celui de « THAM » (Traduction humaine assistée par la machine ») concerne les outils ou environnements d'aide au traducteur ou au réviseur.

Avant de se demander quelle pourrait être la « station de travail du traducteur » en l'an 2001, il n'est pas inutile de voir où nous en sommes aujourd'hui.

TA pour le dépistage

Vers 1949, les États-Unis, puis l'URSS, ont lancé des programmes de « TA » motivés par le besoin de renseignements. C'est ce que nous appellerons la *TAO pour le veilleur.* Il s'agit de traduction automatique, dont on attend des traductions *grossières*, produites rapidement, en grand volume et à bas coût.

La qualité de ces traductions n'est pas essentielle. Elles servent en effet à filtrer des documents, dont les plus intéressants seront, si nécessaire, traduits ou communiqués à des spécialistes bilingues. Préédition et postédition doivent être absentes ou très limitées (ex. : séparer les phrases, les formules, les figures...).

Les systèmes SYSTRAN sont essentiellement de ce type (par exemple, le système russe-anglais installé depuis 20 ans à la Wright-Patterson Air Force Base traduit, d'après nos informations, environ 18 millions de mots par an, avec une qualité tout à fait satisfaisante pour l'usage visé).

Ce besoin est toujours actuel. Cependant, il s'agit maintenant plus de veille scientifique, technique, économique et financière que de renseignement militaire. À titre d'exemple, on peut citer l'accès en anglais à des bases de données japonaises depuis la Suède [4.6].

TA pour la diffusion

Une quinzaine d'années plus tard, on a commencé à travailler sur la *TAO pour le réviseur.* Il s'agit de produire automatiquement des traductions *brutes*, destinées à être révisées. Dans cette optique, la machine doit remplacer le traducteur, qui est promu réviseur. Cela n'est possible que si l'on restreint convenablement le style et le domaine

des textes à traduire (approche par « sous-optimisation », pour reprendre le terme de L. Bourbeau [1.14]).

Les décideurs (politiques, scientifiques et industriels) comme le grand public n'envisagent souvent que cette possibilité, et ce sans doute à tort. En effet, il existe des systèmes qui peuvent répondre à des besoins de ce type, mais seulement dans des situations convenables. Sinon, l'échec est garanti. Voyons cela un peu plus en détail.

Il existe actuellement près d'une quinzaine de systèmes. Il y a surtout des systèmes japonais (AS-Transac de Toshiba, ATLAS-II de Fujitsu, PIVOT de NEC, HICAT de Hitachi, SHALT d'IBM-Japon, PENSÉE de OKI, MU de l'Université de Kyoto et du JICST...), qui traitent presque uniquement les couples japonais <—> anglais. On peut encore citer des systèmes américains (LOGOS, METAL) ou français (Ariane/aéro/F-E de SITE-B'VITAL, fondé sur les outils informatiques et les méthodes linguistiques du GETA) centrés sur l'anglais, l'allemand ou le français, bien que des maquettes ou des prototypes existent sur de nombreuses autres langues.

Que peut-on espérer de tels systèmes? Essentiellement, de répondre à des besoins de plus en plus importants en traduction technique. Typiquement, en moyenne industrielle, une page de 250 mots est traduite en une heure et révisée en 20 minutes. Dans l'idéal, avec quatre personnes, on passerait donc de trois pages à l'heure à douze pages à l'heure, et on multiplierait donc la productivité par quatre. Il s'agit en fait d'une limite, le chiffre le plus raisonnable étant plutôt de huit pages à l'heure, en comptant une révision plus lourde, de 30 minutes par page, et ce avec des réviseurs formés.

Quand les utiliser? Cela n'est actuellement envisageable que pour de gros flux de textes homogènes et informatisés, comme des manuels d'utilisation ou de maintenance. Dans ces conditions, un système à 1 MF (400 KF de base et 600 KF de spécialisation au vocabulaire et au type de texte) doit pouvoir être amorti en deux ans, pour un flux de 10 000 pages par année (en comptant 10 %/an de maintenance, 60 F/page de coût machine, et 100 F/page de révision, contre 150 F/page de traduction et 70 F/page de révision pour la méthode manuelle classique, soit 60 F/page de gain pour amortir 1,2 MF).

Comment les utiliser? Une condition essentielle de succès de ce type de TAO est de constituer une équipe de développement et de maintenance des linguiciels (dictionnaires, grammaires) qui soit en liaison constante avec l'équipe de révision, et si possible avec les auteurs des documents à traduire. C'est ce qu'a su réussir la PAHO (Pan American Health Organization) [4.7], avec ses systèmes ENGSPAN et SPANAM.

Dans le « contre-rapport ALPAC » du JEIDA [1.12] comme au MTS-II à Munich en août 1989 par exemple, Fujitsu a clairement reconnu avoir fait une erreur en distribuant largement ATLAS-II : seules sont en effet rentables les traductions effectuées chez Fujitsu, soit pour sa documentation, soit dans le cadre d'un contrat avec la CEE, ce dernier ne demandant qu'une révision minimale, car il s'agit en fait de veille technologique. Peut-être est-il approprié de faire ici un parallèle avec les systèmes experts, qui peuvent être développés par des tierces parties, mais qui doivent ensuite être totalement maîtrisés par leurs utilisateurs, seuls à même de les faire évoluer de façon adéquate.

Outils pour traducteurs et réviseurs (THAM)

Que faire pour la plus grande partie des textes dont on veut obtenir de bonnes traductions? La bureautique a commencé à apporter des solutions, sous forme d'outils de *TAO pour le traducteur*. Il s'agit ici de traduction assistée, et non plus automatique (THAM, ou Traduction humaine assistée par la machine). C'est l'utilisateur qui traduit, en s'aidant de dictionnaires bilingues, de bases terminologiques, de thésaurus, de *bitextes* (textes + traductions), etc., accessibles depuis un traitement de texte, le tout formant un *poste de travail pour le traducteur*, réalisé sur micro-ordinateur ou station de travail. Il s'agit d'outils comme Mercury/Termex™ sur PC, WinTool™ [6.10] sur Macintosh, ou de systèmes complets (Weidner-Bravice, TSS de Alps). Le réviseur peut utiliser le même environnement, ou préférer travailler directement sous l'outil final de PAO (Publication assistée par ordinateur).

Pour la majorité des besoins, et en particulier pour la traduction de manuels d'enseignement dans des pays où la langue nationale ne s'est que récemment affirmée comme support de l'enseignement secondaire et universitaire, la THAM est actuellement la seule voie réaliste. Il en est de même de toutes les traductions scientifiques et techniques de faible volume, ou hétérogènes, voire de grands volumes homogènes trop mal rédigés (résumés avec des phrases de 15 lignes, par exemple) ou non disponibles sous forme magnétique cohérente et sans erreurs.

Adapter un système de TAO du réviseur à des besoins comparativement ponctuels serait comme réoutiller une usine pour produire quelques dizaines de voitures. En effet, sans compter la saisie optique ou manuelle, entraînant toujours un coût important de vérification, ni la maintenance, ni même l'achat du système de base, mais seulement sa spécialisation (600 KF) et les coûts de traduction et de révision, on arrive à 632, 680, 760 et 920 KF pour 200, 500, 1000 et 2000 pages, contre 44, 110, 220 et 440 KF pour la méthode classique manuelle, soit environ 14,5, 6, 3,5 et 2 fois plus, respectivement. Il faut aussi tenir compte d'un délai de l'ordre de plusieurs mois pour la spécialisation. Tout compte fait, le point d'équilibre se situe à 9000 ou 10 000 pages.

Les « stations » et « environnements » actuels sont-ils adaptés?

Tentons de répondre à cette question en examinant ce qui existe pour la TA-dépistage, la TA-diffusion et la THAM. Nous verrons qu'il conviendra, pour la suite, de nous intéresser d'abord aux utilisateurs, et ensuite aux techniques.

La TA-dépistage

Accès à un serveur

En France, Gachot SA commercialise un serveur de traduction automatique, via le Minitel, qui donne accès à un certain nombre de « paires de langues » de la famille

des systèmes SYSTRAN. Malheureusement, la publicité cherche à faire croire qu'on peut utiliser ce système en thème, pour la diffusion, alors que les résultats sont évidemment désastreux. Ils le seraient d'ailleurs tout autant avec n'importe quel autre système, car aucun traducteur humain, et *a fortiori* aucun système automatique, ne peut traduire n'importe quoi de façon assez convenable pour une diffusion directe, ni même pour une diffusion après révision (car les réviseurs n'acceptent pas de réviser de la trop mauvaise qualité, et préfèrent retraduire, ce qui, au total, est plutôt contreproductif).

Il reste que, pour le dépistage, et donc en version, l'idée du Minitel est remarquable. Il faudrait seulement la compléter en donnant accès à des bases textuelles en langue étrangère. Cette dernière possibilité a été réalisée en Suède [4.6], où l'on peut accéder à des bases documentaires scientifiques et techniques japonaises, via un PC et un modem, la traduction étant encore réalisée par une version japonais-anglais de SYSTRAN.

Enfonçons le clou : pour ce genre de TA, bien que la qualité de la traduction puisse être jugée extrêmement basse par des traducteurs ou des linguistes, les utilisateurs reconstituent assez aisément le sens de ce qu'ils lisent, et c'est la seule chose qui les intéresse.

À la CEE, SYSTRAN est aussi « utilisé » depuis fin 1976. En fait, il a fallu plus de 12 ans aux services de la Commission pour arriver à la conclusion précédente, c'est-à-dire pour renoncer à faire réviser par des professionnels les sorties de SYSTRAN comme des traductions brutes, et pour les proposer directement aux utilisateurs finals comme des traductions grossières. Selon des informations données par M. L. Rolling fin 1990, on est ainsi passé d'un maximum de 2000 pages (traduites automatiquement et révisées) sur 800 000 pages traduites à la CCE (en 1988), soit 2,5 % du total, à près de 40 000 pages (transmises sans révision) sur 900 000 (en 1989), ce qui reste encore faible dans l'absolu (4,4 %), mais montre bien que, pour un tel flux de textes très variés, seule la TA-dépistage est actuellement viable.

En ce qui concerne les « stations de travail » destinées aux réviseurs, on est passé en 12 ans de terminaux 3270 aux mêmes outils que ceux utilisés par les traducteurs, à savoir des micro-ordinateurs Wang, augmentés d'outils de connexion, ainsi que d'utilitaires liés à la saisie optique. Sans doute les lecteurs actuels des traductions grossières utilisent-ils le même matériel, ou leur micro préféré, ou... le papier.

Stations intégrées

Les matériels deviennent assez puissants pour qu'on puisse mettre sur un micro-ordinateur portable un système de TA couplé avec un système de lecture optique, et un éditeur bilingue. C'est ce que proposent plusieurs sociétés japonaises, comme Toshiba et Sharp. Il s'agit tout de même de machines à 15 ou 17 Mips, avec 8 ou 16 Mo de mémoire centrale, et 180 Mo de disque, sous Unix, et coûtant au moins 15 000 $, sans compter le lecteur optique ni le système.

Le résultat est tout à fait impressionnant. On choisit un fragment d'une page (en anglais), on le numérise, on corrige quelques erreurs de saisie, et ça traduit. Là encore,

la traduction japonaise est très loin d'être parfaite, mais suffit, nous dit-on, à comprendre l'original.

Problème de saisie

Il est clair que l'entrée du texte suppose de la part de l'utilisateur une connaissance minimale du système d'écriture de la langue source, ici l'anglais. C'est le cas de tous les utilisateurs potentiels japonais, dont d'ailleurs une bonne partie connaît suffisamment l'anglais pour ne pas avoir besoin de ce type de système. Que se passerait-il pour saisir des textes en russe ou en coréen (sans parler de l'arabe, qui résiste pour l'instant à la lecture optique)? Ou encore, comment inverser ces systèmes, à l'usage d'Occidentaux désireux de lire en anglais des textes japonais?

Il est vrai que la lecture optique du japonais, fondée sur de gros dictionnaires, donne d'excellents résultats, et que celle des écritures alphabétiques (sauf l'écriture arabe et ses dérivés) devrait arriver à des résultats analogues (moins de 1 % d'erreurs sur les mots, et non sur les caractères), en reprenant le même type de technique. Cependant, les meilleures techniques laissent encore une dizaine d'erreurs par page, ce qui, sans correction, diminue considérablement la qualité que peut produire un système de TA.

TA-diffusion

Éditeur adapté sur grand système

Les systèmes de TA tournant en général sur de gros systèmes, on a commencé par adapter les éditeurs standard, en ouvrant automatiquement deux fenêtres (texte source, texte cible), et parfois trois (dictionnaire classique). Les firmes japonaises ont développé des éditeurs spécifiques, où les phrases originales et traduites (qui restent encore aujourd'hui les unités de traduction, sauf dans les systèmes écrits en Ariane!) sont mises en regard.

Depuis l'arrivée de micro-ordinateurs puissants, munis d'éditeurs et de « texteurs » bien plus conviviaux, cette voie est de moins en moins suivie.

Micro-ordinateur lié à un serveur

En général, le réviseur utilise le même environnement que pour de la traduction-révision usuelle, à savoir un micro-ordinateur classique, connecté au serveur de TA.

Par exemple, le système METAL de Siemens tourne sur une machine LISP, la révision se faisant sur un PC de Siemens. De même, les systèmes ATLAS II de Fujitsu, PIVOT de Nec, HICAT de Hitachi, quand ils sont utilisés en interne pour traduire les documents techniques japonais en anglais, tournent sur de grosses machines, la révision se faisant sur des micros connectés.

De même, SITE-B'VITAL fait tourner le système Ariane/aéro/F-E (sous Ariane-G5) sur gros IBM, à partir d'un micro de base (PC sous MS-DOS avec Word), depuis lequel on a aussi accès à la terminologie spécialisée.

Évolution vers l'intégration?

Les matériels deviennent assez puissants pour tout intégrer, comme nous l'avons vu plus haut. Cependant, cette solution ne semble pas la meilleure pour la TA-diffusion, dans la mesure où l'on vise de grands flux, et où la révision doit donc être réalisée par un groupe de traducteurs.

Par exemple, le système METEO [4.3] tourne depuis longtemps sur micro, mais J. Chandioux a conservé l'architecture avec serveur de TA et révision sur des postes individuels.

THAM

Matériel spécifique et un peu de TA : l'échec est garanti!

C'est la THAM qui a créé le concept de « station du traducteur », et... ses dangers. Bien des déboires sont en effet dus à cette idée qu'il faudrait au « traducteur », sans qu'on sache d'ailleurs s'il travaille en indépendant ou en groupe, un environnement matériel et logiciel spécifique.

Ainsi, ALPS et Weidner ont longtemps proposé du matériel spécifique. Celui d'ALPS était d'origine Convergent Technology, tout comme le Questar-400 que SG2 utilisa, avec aussi peu de bonheur, au temps du PN-TAO [4.2], comme station de révision et poste d'indexage pour un système de TA écrit en Ariane-78. Avec ce type de solution, on arrive toujours à des coûts prohibitifs, à des difficultés de connexion, et à une rapide obsolescence technologique, pour la simple raison que les matériels standard, vendus en très grand nombre, peuvent faire l'objet d'investissements très importants, et deviennent plus puissants. C'est aussi, dans un autre domaine, ce qui est arrivé à Xerox avec ses stations de travail d'IA (1165).

L'idée de proposer un éditeur spécifique est tout aussi mauvaise. Au départ, ceux d'ALPS et de Weidner, pour reprendre les mêmes exemples, étaient fort bons. Mais la grande masse des traducteurs, formée d'indépendants, n'a aucune envie d'apprendre un éditeur ou un texteur nouveau, et à coup sûr très vite dépassé par les grands du marché, comme Word, WordPerfect, etc. C'est aussi l'amère expérience qu'a faite A. Melby, avec son premier système TAIM [6.1], intimement lié au texteur Palantir, qui ne connut pas le succès escompté.

Le pire arrive quand on veut ajouter à ces stations de THAM « un peu de TA », comme l'ont fait ALSP et Weidner. En effet, on se trouve contraint à sur-simplifier le traitement linguistique, pour qu'il tourne en temps raisonnable, et surtout pour que

l'utilisateur puisse lui-même modifier les dictionnaires. En 1985, par exemple, le Bureau des traductions du Secrétariat d'État à Ottawa fit une étude sur Weidner, en demandant aux mêmes traducteurs de traduire le même ouvrage (sur l'odontologie), une fois avec la fonction de TA, et une fois sans, avec un intervalle de temps de deux ou trois mois. Avec la TA, on perdit près de 40 % de productivité... Là comme ailleurs, le mieux est l'ennemi du bien.

Ceci expliquant cela, Weidner a fait faillite, et ALPS s'est reconverti en bureau de traduction, et à la sous-traitance pour IBM, avec TSS (*cf. infra*).

Outils de THAM pure : trois cibles

Il semble y avoir actuellement trois tendances en THAM. D'une part, on trouve des *environnements spécifiques destinés à des professionnels travaillant en groupe*, sur de grosses documentations techniques. C'est là que nous classerons le TSS (Translation Support Service) initialement commandé à ALPS par IBM pour VM/CMS (sur grosses machines) et pour OS/2. L'idée essentielle semble d'être d'intégrer la traduction à l'environnement de documentation technique.

En effet, on peut ainsi considérablement réduire les coûts et augmenter la qualité. D'abord, il est très fréquent qu'on traduise de nouvelles versions d'anciens documents, alors qu'une meilleure gestion permettrait de ne retraduire que les passages qui ont effectivement changé. Ensuite, l'accès aux traductions passées peut fournir rapidement la solution à de nombreux problèmes de traduction, et augmenter l'homogénéité de l'ensemble des traductions.

D'autre part, on trouve des *outils destinés aux professionnels indépendants*. Il s'agit surtout de programmes résidant sur PC (comme Mercury/Termex [6.5]) ou d'accessoires de bureau sur Macintosh (comme WinTool), qui permettent d'accéder à un ou plusieurs dictionnaires « en ligne », depuis n'importe quelle application. Il est très important que l'utilisateur puisse définir des dictionnaires personnels et les modifier.

On trouve enfin des *environnements destinés à des traducteurs occcasionnels*, comme SISKEP [6.6, 6.8], déjà bien diffusé en Malaisie. Il s'agit pour l'instant de versions préliminaires, diffusées gratuitement, mais des produits devraient suivre. Là, le traducteur n'est pas un professionnel, et ne traduit pas nécessairement vers sa langue maternelle. Même si c'est le cas, il ne connaît souvent pas certains termes spécifiques, les ayant appris dans la langue source (cas de l'anglais-malais, du français-arabe, par exemple).

La traduction occasionnelle requiert ainsi, outre des dictionnaires bilingues en ligne, des outils liés à la rédaction dans la langue cible, comme des correcteurs d'orthographe, des « critiqueurs » de style, des conjugueurs, des thesaurus, etc.

Peut-être à cause de la simplicité morphologique de l'anglais, SISKEP est le seul système qui effectue une lemmatisation de la chaîne sélectionnée en cas d'échec de la recherche directe dans le dictionnaire, pour éviter à l'utilisateur d'entrer lui-même la forme canonique. Il s'agit d'une propriété très utile, que nous avons cherché à généraliser en concevant notre maquette ODILE [6.9] sur Macintosh.

Propositions, maquettes, idées...

Il y a eu ces dernières années plusieurs études sur la « station du traducteur », avec des maquettes, comme celles du CCRIT au Canada et de CAP-Gemini-Innovation en France. Le sujet est certainement actuel. Ces études renforcent l'idée que les outils de THAM doivent être les moins spécifiques possible, et surtout qu'il faut proscrire tout matériel spécifique.

L'un des problèmes cruciaux pour les traducteurs reste l'absence d'environnements vraiment multilingues. Bien sûr, il existe des adaptations de logiciels à telle ou telle langue, et même, sur le Macintosh, de nombreuses variantes du système d'exploitation, adaptées aux langues de grande diffusion. Il ne s'agit toutefois que de « localisation » : on traduit les messages et les formats particuliers (nombres, dates, prix), et on traite un système d'écriture particulier (arabe, chinois, japonais...), en plus du système de base, fondé sur l'alphabet romain. Mais il reste impossible de fabriquer une nomenclature multilingue comportant, par exemple, du japonais et de l'arabe.

Quatre grandes situations pour le futur

Le concept de « station du traducteur » nous semble trop limité à la traduction professionnelle. Or, les utilisateurs de TAO seront de plus en plus des non-professionnels ou des amateurs. Selon la compétence de l'utilisateur dans les langues visées, et, pour le professionnel, selon qu'il travaille en indépendant ou en groupe, nous pouvons distinguer quatre types de situations pour le futur.

Traduction « dépistage » individuelle

Les serveurs de TA et les solutions intégrées devraient coexister.

Accès à un serveur

Les serveurs paraissent appropriés pour toutes les situations où une information peut être demandée par un grand nombre de personnes. Ainsi, on ne la traduira qu'une fois, et éventuellement à l'avance. Cela permet également d'avoir des versions spécialisées par type de textes, et d'effectuer dans certains cas un « toilettage » très rapide des sorties (comme le fait actuellement Fujitsu sur des traductions grossières fournies à la CCE).

On retrouve ici les situations classiques, où l'information est déjà numérisée : accès à des bases de données documentaires, à des messages éphémères (comme la météo ou les dépêches d'agence). On peut aussi penser à la « lecture active » dans les bibliothèques informatisées du futur.

On peut imaginer un service d'aide à la saisie, dans le cas où le lecteur n'a aucune connaissance du système d'écriture de la langue source et où la lecture optique laisse

à désirer, par manque de qualité de l'original ou toute autre raison. Si l'utilisateur a un lecteur optique, l'aide aurait lieu à distance. Sinon, il se déplacerait ou enverrait l'original.

Solution intégrée sur gros micro

Il s'agit, encore plus que dans le cas précédent, de systèmes à large spectre. On imagine des utilisateurs documentalistes dans de grandes organisations ou firmes, cherchant à savoir ce qui se fait dans tel ou tel domaine, et ce dans plusieurs langues.

Soit l'information est déjà numérisée (disquettes, bandes, DOC, bases de données non munies de serveurs de TA...), et il n'y a « que » des problèmes de traduction de formats. Soit elle ne l'est pas, et l'utilisateur doit acquérir une connaissance minimale du ou des systèmes d'écriture utilisés, ce qui reste tout à fait possible dans le cadre évoqué. Par exemple, il ne faut pas une longue formation pour répondre à un système qui vous demande quel caractère chinois choisir, parmi une liste d'une dizaine.

Vu la tendance actuelle, on risque peu de se tromper en disant qu'il s'agira de vraies « stations de travail », au sens où on l'entend aujourd'hui (10 à 20 Mips, Unix, X-Window).

Nécessité d'environnements génériques (matériel, logiciel)

Dans un cas comme dans l'autre, les environnements seront génériques. La seule différence concernera la puissance requise. Pour l'accès à un serveur, un micro de base suffit et suffira. Pour les solutions intégrées, il faudra une station de travail standard.

Dans le premier cas, il est évident qu'on ne peut imposer un logiciel spécifique pour l'édition des résultats. Il faudra fournir des « filtres » (transformateurs de formats) pour les standards du marché. Dans le second cas, on ne le peut pas non plus, car il est fort probable que l'utilisateur devra mettre plusieurs systèmes sur sa station, ayant à traduire à partir de plusieurs langues sources. En effet, pour les constructeurs de systèmes de TA, il est plus rentable et plus facile de travailler à partir d'une ou deux langues vers de nombreuses langues que l'inverse.

Traduction « occasionnelle »

Il s'agit de personnes qui connaissent bien les deux langues, mais pas comme des traducteurs, et qui ne traduisent pas nécessairement vers leur langue maternelle, ni même à partir d'elle. En Malaisie, par exemple, il est fréquent que des Chinois rédigent des notes de cours en anglais et les traduisent ensuite en malais. En Afrique du Nord, de nombreux scientifiques formés en France peuvent avoir à traduire vers l'arabe à partir de l'anglais aussi bien que du français.

Dans ce contexte, il faut absolument des outils simples, tournant sur des micros de bas de gamme, si possible en conjonction avec la plupart des applications du commerce (tableurs, texteurs, grapheurs, SGBD, documenteurs...).

Étant donné les progrès techniques et la baisse rapide des prix, on peut penser trouver, sur de tels matériels, les fonctions actuelles (recherche dans les dictionnaires bilingues avec lemmatisation automatique, thesaurus en langue cible, fonctions de manipulation morphologique en langue cible), présentées de façon bien plus ergonomique, c'est-à-dire au moyen de fenêtres « vocables » qui seraient liées à la fenêtre de l'application (texteur par exemple), et contiendraient en permanence les informations associées aux termes de cette fenêtre. Le travail de recherche se faisant en tâche de fond, l'utilisateur n'aurait plus à sélectionner, demander une recherche, attendre, etc.

Traduction professionnelle « individuelle »

Le traducteur professionnel traduit en principe dans sa langue, ou dans une langue qu'il connaît parfaitement, ainsi que la plupart des termes, généraux et techniques, de ce qu'il doit traduire. Par conséquent, si les outils précédents peuvent lui être utiles, des outils de recherche intelligente dans ses traductions passées lui seraient bien plus profitables.

Des outils de communication sont également indispensables, pour la consultation de bases terminologiques dans les cas particulièrement délicats. Même si on a un DOC, aucune base terminologique n'est jamais complète.

Par conséquent, on voit bien un traducteur indépendant s'équiper d'un micro « classique », plutôt de haut de gamme (AT-386 sous Windows, ou Mac LC, par exemple), avec beaucoup de mémoire sur disque, un lecteur de DOC et des outils de connexion.

Verra-t-on des éditeurs ou texteurs spécialement conçus pour les traducteurs indépendants? C'est fort douteux, vu que ce marché est étroit, que les clients demandent les formats usuels, et que les texteurs offriront sans doute la plus grande partie des fonctionnalités souhaitables, comme la possibilité de synchroniser deux ou plusieurs fenêtres (comme WinText le fait déjà).

Traduction professionnelle « industrielle »

La traduction professionnelle de haute qualité est souvent produite par plusieurs personnes. Habituellement, le traducteur qui effectue le premier jet a une compétence technique très superficielle, mais connaît très bien la terminologie et les deux langues. Le premier réviseur est un traducteur « senior », spécialiste du type de document considéré, et à même d'assurer l'homogénéité terminologique et stylistique. Enfin, on fait parfois intervenir un second réviseur, spécialiste du contenu technique du document et éventuellement ignorant de la langue source, pour détecter les contresens sémantiques

linguistiquement plausibles, ainsi que d'éventuelles ambiguïtés accidentellement introduites en langue cible.

D'autre part, les documents à traduire sont souvent très longs, et doivent être traduits très rapidement. Il faut donc synchroniser le travail de plusieurs traducteurs et réviseurs. Ainsi, la traduction professionnelle industrielle justifie-t-elle la conception d'environnements plus spécifiques que la traduction professionnelle individuelle.

L'introduction de la TA dans ce type d'activité se généralisera sans doute. Aux fonctionnalités mentionnées ci-dessus, il faudra donc rajouter celles qui ont trait à l'amélioration du système de TA. Pour cela, il faut pouvoir recenser les corrections des réviseurs, leur permettre d'envoyer des messages aux développeurs-mainteneurs du système, et éventuellement leur donner accès, sous une forme appropriée, aux lexiques du système, pour leur permettre de tenter un diagnostic, et en tout état de cause supprimer l'aspect « boîte noire », psychologiquement très néfaste.

La fonction de TA étant réalisée par un serveur, pour les raisons indiquées plus haut, il n'est pas nécessaire de viser des stations de travail. Le type de matériel nécessaire sera plutôt fonction de la puissance requise par l'ensemble des outils liés à la gestion de la documentation technique et de la traduction, s'il s'agit de micros. Mais il pourra tout aussi bien s'agir de « terminaux intelligents », successeurs des « terminaux X » actuels, connectés à un serveur de gestion de la documentation, de la traduction et des tâches de TA.

Perspectives et conclusions

Nous avons évoqué quatre types d'utilisateurs, le veilleur, le traducteur occasionnel, le traducteur professionnel indépendant et le traducteur professionnel industriel. Seuls le premier et le quatrième peuvent à notre avis utiliser avec profit les techniques de TA existantes. Il reste encore une possibilité, actuellement au stade des recherches préliminaires, celle de la TAO personnelle [5.8] ou *TAO pour le rédacteur.*

De plus en plus, nous désirons rédiger dans notre langue, et transmettre nos textes à l'étranger, qu'il s'agisse de messages électroniques, de lettres, d'articles (on pense à la récente controverse sur l'Institut Pasteur), de manuels techniques, voire de livres. En effet, contrairement à ce que d'aucuns prédisaient il y a une cinquantaine d'années, l'internationalisation croissante ne s'est pas accompagnée d'une uniformisation linguistique vers l'anglais, mais au contraire d'un renforcement considérable de l'usage scientifique et technique des langues traditionnellement importantes de ce point de vue, et d'une promotion volontariste de bien d'autres, pour les amener au même niveau (malais-indonésien ou arabe, par exemple). À notre sens, cette évolution ne fera que se renforcer, les langues étant, comme le note Claude Hagège dans un article paru dans *Le Monde* début 1990, les *drapeaux des identités nationales.*

Au GETA, nous sommes en train d'étudier et de maquetter ce nouveau concept de « TAO personnelle » dans le cadre du projet LIDIA. Une hypothèse de base est que le dialogue nécessaire, même s'il est lourd (de l'ordre de grandeur de la traduction-révision humaine vers une seule langue cible), serait acceptable dans la mesure où

l'auteur, et non le système, le déclencherait, où une réalisation asynchrone éliminerait toute attente forcée, et où le système serait « contrôlable » et « découvrable » par l'utilisateur, grâce à des possibilités de réglage de paramètres, d'intervention au niveau des « marques cachées », de « rétrotraduction », et d'exploration des connaissances lexicales et grammaticales contenues dans le système, présentées de façon « naturelle » et non codée, ce qui est d'ailleurs un problème en soi.

Les trois types de TAO *classique* ne conviennent évidemment pas à ce nouveau besoin. En effet, la TAO du veilleur (TA-dépistage), sans préédition ni postédition, ne peut donner une qualité suffisante, et la TAO du réviseur (TA-diffusion) comme la TAO du traducteur (THAM) s'adressent par définition à des spécialistes au moins bilingues, et non à des rédacteurs du grand public, supposés ne connaître aucune des langues cibles (ou au plus une, et ce imparfaitement).

La « station du rédacteur », dans un système de TAO personnelle, devrait, pour des raisons de diffusion, être un micro-ordinateur répandu, de bas ou de milieu de gamme. Elle supporterait les logiciels de vérification orthographique, terminologique et stylistique, de dialogue, et de communication, tandis que le système de TA tournerait sur un serveur (connecté à un réseau local ou accessible par Minitel). Du point de vue logiciel, il faudra utiliser un outil programmable du marché. Aucun texteur ne l'étant pour l'instant, nous avons choisi de maquetter avec HyperCard, qui présente d'ailleurs d'autres avantages, tant ergonomiques que linguistiques [5.5]. Mais cela évoluera sans doute. En tout état de cause, il y a encore beaucoup de recherche à faire sur ce concept, puis d'importants investissements linguistiques, avant de voir des produits de TAO personnelle. Espérons que ce sera une perspective proche en l'an 2001 !

Références (thématiques et chronologiques)

1. Références générales sur la TAO

[1.1] KITTREDGE R. (1983) « Sublanguage — Specific Computer Aids to Translation — a Survey of the Most Promising Application Areas », *Contract* n° 2-5273, Montréal, Université de Montréal et Bureau des traductions, mars 1983, 95 p.

[1.2] BOITET Ch. (1985) *Traduction (assistée) par ordinateur : ingénierie logicielle et linguicielle*, Conf. AFCET RF&IA, Grenoble, nov. 1985.

[1.3] HUTCHINS W. J. (1986) *Machine Translation : Past, Present, Future*, Chichester, England, Ellis Horwood, John Wiley & Sons, 382 p.

[1.4] KITTREDGE R. (1986) *Analyzing Language in Restricted Domains : Sublanguage Description and Processing*, Grishman R. & Kittredge R. (eds.), Hillsdale, New-Jersey, Lawrence Erlbaum.

[1.5] DESCLÉS J. P. (1987) *Technologos*, Paris, LISH-CNRS, printemps 1987.

[1.6] BOITET Ch. (1988) « Software and Lingware Engineering in Modern M(A)T Systems », Bátori (ed.) *Handbook for Machine Translation*, Tübingen, Niemeyer.

[1.7] BOITET Ch. (1988) « PROS and CONS of the Pivot and Transfer Approaches in Multilingual Machine Translation », *New Directions in MT*, BSO conf, Budapest, 13 p.

[1.8] BOITET Ch. (1988) *L'apport de Bernard Vauquois à la traduction automatique et au traitement automatique des langues naturelles*, Colloque sur l'Histoire de l'informatique en France, Grenoble, 3-5 mai 1988, vol. 2, p. 63-82.

[1.9] LEHRBERGER J. et BOURBEAU L. (1988) « Machine Translation. Linguistic Characteristics of MT Systems and General Methodology of Evaluation », The Hague, John Benjamins, *Lingvisticæ Investigationes*, 15, 240 p.

[1.10] VAUQUOIS B. (1988) *BERNARD VAUQUOIS et la TAO, vingt-cinq ans de traduction automatique, ANALECTES. BERNARD VAUQUOIS and MT, twenty-five years of MT*, Ch. Boitet (éd.), Grenoble, Ass. Champollion & GETA, 700 p.

[1.11] ABBOU A., (éd.) (1989) *Traduction assistée par ordinateur*, Actes du séminaire international (Paris, 17-18 mars 1988) et dossiers complémentaires, Paris, DAICADIF, Observatoire des Industries de la langue, 234 p.

[1.12] JEIDA (1989) *A Japanese View of Machine Translation in Light of the Considerations and Recommendations Reported by ALPAC, USA*, Tokyo, Japanese Electronic Industry Development Association, 197 p.

[1.13] BOITET Ch. (1990) « Multilingual Machine Translation Does Not Have to Be Saved by Interlingua », *Proc. MMT'90*, Tokyo, 5-6 novembre 1990.

[1.14] BOURBEAU L. (1990) *Élaboration et mise au point d'une méthodologie d'évaluation linguistique de systèmes de Traduction Assistée par Ordinateur*, rapport final, Ottawa, Secrétariat d'État du Canada, mars 1990, 203 p.

[1.15] ROLLING L. (1990) « Trends of Multilingual Machine Translation in Europe », *Proc. MMT'90*, Tokyo, 5-6 novembre 1990, 2 p.

2. Compréhension et traduction (humaine, automatique)

[2.1] BAR-HILLEL Y. (1960) « The Present Status of Automatic Translation of Languages », F.L. Alt (ed.), *Advances in Computers (volume 1)*, New York, Academic Press, p. 91-163.

[2.2] BAR-HILLEL Y. (1964) *Language & Information*, Addison Wesley.

[2.3] BAR-HILLEL Y. (1971) « Some Reflections on the Present Outlook for High Quality Machine Translation », *Feasibility Study on Fully Automatic High Quality Machine Translation*, RADC-TR-71-295, Austin, Univ. of Texas.

[2.4] NIRENBURG S. et GOODMAN K. (1990) « Treatment of Meaning in MT Systems », *Proc. ROCLing-III*, Taipeh, 20-22 août 1990, p. 83-101.

3. TA « fondée sur le contenu »

[3.1] WILKS Y. (1973) « An Artificial Intelligence approach to Machine Translation », Shank & Colby (eds.), *Computer Models of Thought and Language*, Freeman & Co, p. 114-151.

[3.2] BOITET Ch. et GERBER R. (1986) « Expert Systems and Other New Techniques in MT », Bátori & Weber (eds.), *Neue Ansätze in maschineller Sprachübersetzung*, Tübingen, Niemeyer, p. 103-119.

[3.3] BOITET Ch. (1988) « Representation and Computation of Units of Translation for Machine Interpretation of Spoken Texts », *TR-I-0035*, ATR, Osaka, 41 p. Also *Computers & Artificial Intelligence 8 /6*, p. 505-546, 1989.

[3.4] NIRENBURG *et al.* (1989) *KBMT-89 Project Report*, Pittsburgh, Center for Machine Translation, Carnegie Mellon University, February 1989, 286 p.

[3.5] NIRENBURG S. (1989) « Knowledge-Based Machine Translation », *Machine Translation **4/1***, March 1989, 5-24.

[3.6] NIRENBURG S. et DEFRISE Ch. (1990) « Lexical and Conceptual Structure for Knowledge-Based Machine Translation », *Proc. ROCLing-III*, Taipeh, 20-22 août 1990, p. 105-130.

[3.7] CARLSON L. et NIRENBURG S. (1991) « World Modeling for NLP », *CMT*, *CMU*, 31 p.

[3.8] BOITET Ch. (1991) « Un système de traduction automatique peut-il et doit-il comprendre? » Paris, Actes de la Convention IA-91, 15-17 janvier 1991, *Hermès*, p. 13-25.

4. TA « fondée sur la forme »

[4.1] BOITET Ch. et NEDOBÉJKINE N. (1981) « Recent Developments in Russian-French Machine Translation at Grenoble », *Linguistics **19** (1981)*, p. 199-271.

[4.2] BOITET Ch. (1986) « The French National MT Project : Technical Organization and Translation Results of CALLIOPE-AERO », *Computers and Translation **1*** (1986), p. 281-309.

[4.3] CHANDIOUX J. (1988) « 10 ans de METEO CMD) », *In [1.11, Abbou 89]*, p. 169-173.

[4.4] DUCROT J. M. (1988) « Le système TITUS IV », *In [1.11, Abbou 89]*, p. 55-71.

[4.5] PECCOUD F. (1988) « The Aims of the French National Project of Computer-Aided Translation », *International Forum on Information and Documentation*, **13**/1, p. 11-13.

[4.6] SIGURDSON J. et GREATEX R. (1987) « MT of On-Line Searches in Japanese Data Bases », *RPI*, Lund Univ., 124 p.

[4.7] VASCONCELLOS M. et LEÓN M. (1988) « SPANAM and ENGSPAM : Machine Translation at the Pan American Health Organization », J. Slocum (ed.), *Machine Translation Systems*, Cambridge Univ. Press, p. 187-236.

5. TAO « fondée sur le dialogue »

[5.1] RICHARDSON S. D. (1985) *Enhanced Text Critiquing using a Natural Language Parser : the CRITIQUE System*, Yorktown Heights IBM Research Report RC 11332.

[5.2] CHANDLER B., HOLDEN N., HORSFALL H., POLLARD E. et MCGEE WOOD M. (1987) *N-Tran Final Report*, Manchester, Alvey Project, CCL/UMIST Report 87/9, 30 p.

[5.3] ZAJAC R. (1988) « Interactive Translation : a New Approach », *Proc. COLING-88*, Budapest, 22-27 août 1988, p. 785-790.

[5.4] BOITET Ch. (1989) « Motivations and Architecture of the LIDIA Project », *Proc. MTS-89*, Munich, août 1989, 5 p.

[5.5] BOITET Ch. (1989) « TAO Personnelle et HyperTexte », *Actes du colloque BUROTICA-89*, Paris, 10-13 octobre 1989, 6 p.

[5.6] BOITET Ch. (1989) « Speech Synthesis and Dialogue-Based Machine Translation », *Proc. ASTI Conf.*, ATR, Nara, 11-12 décembre 1989, 9 p.

[5.7] BLANCHON H. (1990) *LIDIA-1 : un prototype de TAO personnelle pour rédacteur unilingue*, X^{e} journées internationales sur « les systèmes experts et leurs applications », Avignon, juin 1990, 10 p.

[5.8] BOITET Ch. (1990) « Towards Personal MT : General Design, Dialogue Structure, Potential Role of Speech », *Proc. COLING-90*, Helsinki, 20-25 août 1990, 6 p.

[5.9] BLANC E. & BOITET Ch., (eds) (1990) « DBMT-90, a Workshop on Dialogue-Based Machine Translation : Documents & Slides », *Le Sappey*, 26-28 août 1990, Grenoble, GETA, IMAG, 350 p.

[5.10] BOITET Ch. (1990) « TAO personnelle et promotion des langues nationales : le projet LIDIA du GETA », *Tribune des Industries de la Langue*, Grenoble, automne 1990.

6. Outils de THAM

[6.1] MELBY A. K. (1978) « Design and Implementation of a Computer-Assisted Translation System », *Proc. COLING-78*, Bergen, 14-18 août 1978, 28 p.

[6.2] HUNDT M. G.. (1982) « Working With the Weidner Machine-Aided Translation System », V. Lawson (ed.) *Practical Experience of Machine Translation*, The Hague, ASLIB, North Holland, Conf., p. 45-51.

[6.3] MELBY A. K. (1982) « Multilevel Translation Aids in a Distributed System », *Proc. COLING-82*, Prague, 5-10 juillet, North Holland, p. 215-220.

[6.4] LONSDALE D. (1984) *Notes on Interactive Translation*, ALPS Inc., Provo, Utah, 10 p.

[6.5] MELBY A. K. (1984) *MERCURY ™ Card File and Glossary Manager. User's Guide*, Provo, Utah, LinguaTech Inc., 114 p.

[6.6] TONG L. C. (1986) *SISKEP, a Software for the Human Translator*, PTMK, Penang, Universiti Sains Malaysia, 6 p.

[6.7] KJAERSGAARD P. S. (1987) « REFTEX, un progiciel pour la traduction assistée par ordinateur », *Computers and Translation*, 7 p.

[6.8] TONG L. C. (1987) « The Engineering of a Translator Workstation », *Computers and Translation 2/4*, 263-273.

[6.9] TOMASINO I. (1990) *ODILE, un outil d'intégration extensible de lemmatiseurs et de dictionnaires*, Mémoire CNAM, GETA, Grenoble, décembre 1990, 151 p.

[6.10] WINSOFT (1987) *Manuel d'utilisation de WinTool ™. Version 1.1*, Grenoble, WinSoft.

2

L'environnement sémantique, aide au transfert lexical en TA. Problèmes posés par le calcul des traductions françaises de quelques verbes anglais

Arlette ATTALI, Annette EUVRARD

CELTA, CNRS/EUROTRA F-Nancy, Université de Nancy II, Nancy, France

Nous nous proposons de présenter une synthèse d'un travail réalisé en collaboration avec des membres de l'équipe du Centre d'études linguistiques pour la traduction automatique.

Cette étude s'intègre dans le cadre général des problèmes rencontrés dans tout programme de TA, lequel comporte, nous le rappelons, trois phases : l'**analyse**, le **transfert** et la **génération**. Nous nous sommes plus particulièrement intéressés aux difficultés propres à la **phase transfert**; celle-ci étant le moment où le signifiant de la langue cible se substitue au signifiant de la langue source.

Dans le cas qui nous occupe, c'est-à-dire la traduction bilingue anglais → français, nous avons cherché des moyens linguistiques susceptibles de nous aider à résoudre les problèmes spécifiques liés au **transfert lexical des verbes anglais** en TA[1].

1. *Cf.* publication : « Transfert lexical et contexte sémantique en Traduction Automatique. Les traductions françaises des verbes anglais DRAW, HOLD, LIE, TURN, WORK ». A. Attali, G. Bourquin, M. C. Bourquin-Launay, A. Euvard. CELTA, URA-CNRS 1035. Cahiers du CRAL, Nancy, 1989.

Pour mieux comprendre ces problèmes, nous prenons l'exemple du verbe HOLD, appréhendé dans des textes réels et limités à la phrase, avec une multiplicité de traductions possibles (*cf.* annexe 1). Se pose donc la question de savoir comment un automate-traducteur qui traduit de lexème à lexème peut sélectionner l'équivalent lexical adéquat : chaque traduction française correspondant à un des sens de HOLD doit être l'**aboutissement d'un calcul**. Ce calcul consiste, pour nous, à tirer parti de l'**environnement syntactico-sémantique** porté par les **actants** du verbe anglais. Par actants nous entendons : les sujets profonds ou GN1, les objets directs ou indirects profonds ou GN2 et parfois GN3 auxquels on pourrait ajouter des circonstants plus ou moins attendus.

Le principe est donc de dégager de cet environnement actanciel des informations **minimales** mais aussi **représentatives d'un nombre suffisamment élevé d'exemples** pour pouvoir servir de critères de sélection des traductions dans un programme automatique.

Étant donné la rigidité du système, on est confronté à une double difficulté avec laquelle il faut composer : 1) difficulté du choix de l'équivalent lexical français qui doit couvrir le plus grand nombre de cas possibles et en même temps préserver le sens originel, et 2) difficulté du choix des critères de calcul de cet équivalent lexical, les deux choix étant liés l'un à l'autre.

Nous avons cherché un début de réponse à ces problèmes en commençant par examiner **cinq verbes anglais** : DRAW, HOLD, LIE, TURN, WORK, dans toutes leurs formes grammaticales, à tous leurs temps et accompagnés de toutes leurs particules. Ils ont été relevés dans un corpus de concordances de 3 500 000 occurrences parmi lesquelles ces verbes représentent 5465 attestations.

C'est cette quantité d'exemples tirés de textes authentiques — et non fabriqués — sur laquelle s'appuie notre étude, qui nous permet de répertorier certaines régularités contextuelles afin d'élaborer les règles de calcul applicables au transfert lexical des verbes.

– Nous avons choisi ces cinq verbes pour les raisons suivantes :

1) Ils sont sémantiquement différents les uns des autres.
2) Chacun d'eux a des effets de sens très différents en anglais : concret/abstrait, propre/figuré.
3) Chaque verbe donne lieu à un nombre assez élevé de traductions françaises différentes.
4) Chacun a une fréquence d'emploi suffisamment élevée dans les corpus pour que l'étude soit significative.
5) Enfin, chacun s'intègre dans des réseaux actanciels suffisamment diversifiés et riches.

– La mise en œuvre de l'expérience se décompose en une suite de tâches strictement hiérarchisées. Chaque verbe est examiné cas par cas dans les phrases où il apparaît. Il faut donc :

1) Choisir **empiriquement** la traduction française en fonction du contexte; cette traduction se veut aussi exacte et généralisable que possible.

2) Déterminer le minimum d'information syntactico-sémantique nécessaire au calcul de cette traduction.
3) Élaborer des règles informatisables qui représentent et codifient l'information jugée pertinente.
4) Construire l'algorithme qui rend compte du déroulement hiérarchisé des règles de calcul et qui est représenté sous forme d'organigramme.

Ces différentes étapes se conditionnent l'une l'autre et c'est par approximations successives qu'on parvient finalement au choix de traduction.

Rappelons que l'élaboration de ces règles de traduction suppose qu'un certain nombre de problèmes aient été déjà résolus au cours de la phase d'analyse qui précède, sur le plan syntaxique et sémantique. Par exemple, on considère comme acquises les levées d'ambiguïtés grammaticales (substantif/verbe; adjectif/adverbe); la reconnaissance des référents des pronoms relatifs et personnels, celle des circonstants non prépositionnels et des différentes fonctions syntaxiques en général.

Critères de sélection des traductions fondés sur la morpho-syntaxe

Nous ne donnerons que quelques exemples indicatifs :

1- **Un mot spécifique**; par exemple la particule qui accompagne le verbe :
   – Field **DREW AWAY** on his motorcycle.
   Règle : DRAW + AWAY ⟶ « S'ÉLOIGNER ».

2- **Un élément grammatical** :
   – The observation **HOLDS true** whether...
   Règle : HOLD + Adj. ⟶ « RESTER ».

3- **La voix** à laquelle le verbe est lui-même employé peut être un élément qui oriente vers la bonne traduction :
   – I **HELD** receptions for leading figures
   Règle : HOLD = ACTIF ⟶ « ORGANISER ». (Ici ce n'est pas le seul élément à être pris en compte mais il est nécessaire.)

4- **Une fonction syntaxique** :
   – The culture was **HELD at 34 degrees F**.
   Règle : HOLD + circ. de mesure ⟶ « MAINTENIR ».

Critères de sélection des traductions fondés sur la sémantique :

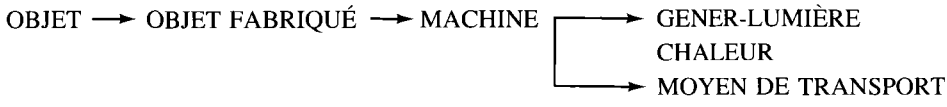
C'est ici qu'interviennent les **descripteurs sémantico-cognitifs** que l'on peut définir comme classes notionnelles regroupant un ensemble d'unités lexicales ayant un trait sémantique commun. Ils se caractérisent de plusieurs manières :

Ces descripteurs ne sont ni de même niveau ni de même nature.

En effet, comme ils ont été établis de manière empirique, **en fonction de la traduction souhaitée**, il y a une grande disparité dans l'étendue de leur champ d'application respectif.

À côté des grands traits sémantiques généraux bien connus tels que HUMAIN, ANIMAL, ABSTRAIT, OBJET, etc., d'autres font référence à des notions très spécifiques : AFFECTIVITÉ, FINANCE, CONNOTATION PÉJORATIVE, PORTEUR D'INFORMATION, PORTEUR DE VIRTUALITÉS, etc. (*cf.* annexe 2).

Ces descripteurs ne sont pas hiérarchisables, tout au plus peut-on en regrouper certains en allant du plus général au plus particulier :



Il convient de souligner que le descripteur ne rend pas compte de la totalité du sémantisme de l'unité lexicale à laquelle il est affecté; il ne retient qu'un des aspects couvrant un champ plus ou moins large. C'est ainsi que certains lexèmes peuvent être classés sous plusieurs descripteurs différents. Un cas intéressant est par exemple le mot « **POSITION** » classé sous quatre descripteurs différents et « **BOOK** » sous trois (*cf.* annexe 3).

Ces descripteurs sont de rentabilité variable.

Un descripteur est « rentable » lorsqu'il est fréquemment utilisé : soit parce qu'il recouvre un grand nombre de termes, soit parce qu'il est utilisé pour plusieurs verbes, soit parce qu'il réalise les deux choses à la fois.

C'est le cas pour le descripteur **HUMAIN**, statistiquement le plus rentable : il est utilisé pour les cinq verbes, il recouvre un très grand nombre de mots et il est appliqué aux trois fonctions GN1, GN2, GN3.

Par contre le descripteur « **PHÉNOMÈNE SOCIAL TEMPORAIRE** » ne concerne que le verbe HOLD, mais s'applique à un très grand nombre de termes. De plus il est sémantiquement bien délimitable (*cf.* annexe 4).

Le descripteur « **ENTITÉ SPATIALE, GÉOGRAPHIQUE, GÉOLOGIQUE** » concerne trois verbes: DRAW, HOLD et LIE; mais comme il est très général, la limite de son extension est difficile à établir (*cf.* annexe 4).

Enfin, on a créé des descripteurs, pour le moment très faiblement rentables, comme « **TEMPS** », « **BRUIT** », « **PHÉNOMÈNE NATUREL** » (tels que « WAVE LIFT ») qui ne concernent chacun qu'un seul verbe, ne recouvrent qu'un seul mot lui-même attesté dans un seul exemple. Cela est une exception, car nous n'avons pas voulu faire des traductions *ad hoc*; mais dans ces cas on a pensé que

ces descripteurs étaient susceptibles d'accroître leur rendement au fur et à mesure qu'augmenterait le nombre des verbes étudiés. L'expérience montrera si cette intuition était fondée ou non.

Application des descripteurs aux actants concernés :

Les exemples suivants (tirés de l'annexe 1) mettent en évidence le rôle de la sémantique actancielle pour aider au transfert lexical des verbes dans un programme de TA.

1. The **senate seat** *held* by her husband Richard.
   GN2
   [socio-prof. abstrait] **occuper**
4. The century **dinner** *was held* at Eton.
   GN2
   [phén. social temp.] **avoir lieu**
11. The **incubator** *holds* yoghourt jars.
    GN1
    [contenant] **contenir**
14. Party **leaders** *holding* **shares** in private firms...
    GN1 **GN2**
    [humain] **détenir** **[finance]**
15. A surplus of labor *holds back* the **mechanization** of...
    GN2
    freiner **[processus]**

Limites liées au choix des descripteurs sémantiques :

Dans l'idéal, les règles édictées par l'algorithme de traduction devraient permettre de traduire avec exactitude la totalité des attestations du corpus utilisé. Mais ce n'est pas le cas.

Aussi à côté des traductions jugées « **satisfaisantes** » (qui représentent tout de même environ 85 % des cas dans notre corpus), nous avons des traductions dites « **approximatives** » (7 % — il s'agit des cas où la plus grande partie du sens source est conservée), des traductions « **impropres** » (3 % — cas où la plus grande partie du sens source est perdu, mais reste compréhensible grâce au contexte), et enfin des traductions « **incorrectes** » (3 % — cas où l'on a perdu totalement et le sens source et la compréhension).

Nous allons illustrer les TROIS derniers cas (c'est-à-dire les traductions non satisfaisantes) en essayant d'analyser les causes d'imprécisions ou d'erreurs.

– **Première cause** de traductions non satisfaisantes : le fait que nous n'avons pas voulu produire des descripteurs entièrement *ad hoc* afin de ne pas multiplier les traductions uniques (nous ne l'avons fait que pour trois ou quatre cas bien particuliers dont nous avons parlé plus haut). Cela entraîne une traduction par défaut, inadaptée.

Exemple :

The relation theory does not **HOLD UP** very well.

Règle suivie : HOLD + UP intr. → « SE MAINTENIR » au lieu de « RÉSISTER ».

Il aurait fallu créer un trait sémantique sous le GN1 « theory » pour ce seul cas.

Autre exemple :

A wide-ranging force [of Republicans]... was **DRAWN UP** against a solid phalanx of southern Democrats.

Règle suivie : DRAW + UP + GN2 → traduction par défaut → « ÉTABLIR ».

« METTRE SUR PIED » eût été meilleur. Mais là aussi, il aurait fallu créer un descripteur pour le GN2 : « force » seul cas d'espèce.

Dans ces deux exemples on a affaire à des traductions « **approximatives** » : la plus grande partie du sens est conservée.

– **Deuxième cause** de traduction non satisfaisante : le descripteur est **trop général**.

Exemple :

Very few if any among them HELD SLAVES.

Règle suivie : HOLD + GN2;

si GN2 = « HUMAIN » et GN1 = « HUMAIN » → « TENIR ».

Ici le descripteur « HUMAIN » sur le GN2 **slave** est trop vaste; il aurait fallu ajouter un trait comme « infériorité » ou « bien possédé » pour pouvoir traduire par « AVOIR » ou « POSSÉDER ». Mais là encore c'est le seul cas rencontré.

Ici la traduction est « impropre » : le sens source est perdu mais cela reste compréhensible.

– **Troisième cause** de traduction non satisfaisante : le **caractère limité** de nos investigations qui ne portent que sur la phrase et donc sur un environnement trop étroit.

Exemple :

Sometimes he **DREW** birds on his copy-book.

Règle : DRAW + GN2; si GN2 = « ANIMAL » → « ATTIRER » au lieu de « DESSINER ».

Il aurait fallu pouvoir détecter une information sur le circonstant et la mettre en relation avec les autres éléments. Il faudrait faire appel à des techniques d'IA pour formaliser un univers textuel plus vaste.

Ici la traduction est totalement **incorrecte** : perte de sens.

– **Dernière cause** : **l'ambiguïté** totale du verbe dans la phrase. Dans ce cas le système automatique sort deux traductions possibles et c'est au réviseur humain à choisir.

Exemple : Steve **DREW** a glass of water.

Il est impossible de choisir entre « TIRER » ou « DESSINER » car ici aussi le contexte est trop étroit; il faudrait prendre en compte les informations données antérieurement dans le récit.

Conclusion

À partir de cette étude, limitée alors à cinq verbes, on peut constater l'aide effective apportée par l'environnement syntactico-sémantique au transfert lexical en TA. Les premiers résultats obtenus par une recherche empirique peuvent donner l'impression de manquer de bases solides; c'est pourquoi il nous apparaît que tout progrès dans l'automatisation de la démarche de traduction exige des recherches plus poussées sur un contexte plus élargi, notamment en direction de l'IA.

Quant aux descripteurs sémantiques que nous avons élaborés jusqu'ici, ils sont appelés à évoluer en fonction de l'extension de notre étude à d'autres verbes. Certains descripteurs disparaîtront, d'autres fusionneront, d'autres seront créés et d'autres se confirmeront. À ce propos, par exemple, nous avons constaté que le descripteur « PORTEUR D'INFORMATION » déjà productif pour DRAW, HOLD et TURN, se révèle à nouveau utile pour certaines traductions de RUN et MOVE étudiés récemment.

Enfin, on peut dire que le fait **d'extraire les seuls éléments sémantiques indispensables** pour le choix de traductions dans nos types de textes, est préférable à l'application d'une théorie sémantique préétablie.

Annexe 1

1. The senate seat *held* by her husband Richard.
   occuper
2. He *held* a reception.
   organiser
3. The transplant surgeon *was held* the immediate cause of the death.
   tenir... pour
4. The century dinner *was held* at Eaton.
   avoir lieu
5. ...the man who *holds* the record
   détenir
6. To *hold* a herd of cattle on the range.
   garder
7. Rosemary's mother was *holding out* her long hand.
   tendre
8. Research workers *hold* an opposite opinion.
   soutenir
9. He *was holding* a carabin.
   tenir
10. Inventions are *held up* time and again.
    retarder
11. The incubator *holds* yoghourt jars.
    contenir

12. Joe *held up* the book again.
présenter
13. The technique *holds* great promise...
contenir
14. Party leaders *holding* shares in private firms...
détenir
15. A surplus of labor *holds back* the mechanization of agriculture.
freiner

Annexe 2

AFFECTIVITÉ	**FINANCE**
admiration	benefit
fury	dollar
glamor	money
hatred	salary
suspicion...	share...

PORTEUR D'INFORMATION	**PORTEUR DE VIRTUALITÉS**
data	future
idea	hope
letter	menace
page	possibility
problem...	promise...

Annexe 3

position	entité spatiale (draw, hold, lie)
position	lieu (hold, work)
position	opinion (hold)
position	socioprofessionnel abstrait (hold)
book	contenant (hold, turn)
book	objet (draw, hold, turn)
book	porteur d'information (draw, hold, turn)

Annexe 4

PHÉNOMÈNE	**SOCIAL**	**TEMPORAIRE**
census	debate	meeting
ceremony	dinner	service
congress	election	show
course	festival	trial...

ENTITÉ	**SPATIALE**	**et GÉOGRAPHIQUE**
area	hill	place
city	lake	pool
field	land	sea
glacier	park	valley...

3

Problèmes de désambiguïsation interactive et TAO personnelle

Hervé BLANCHON

GETA, Institut IMAG (UJF & CNRS), Grenoble, France

Résumé

Nous nous plaçons ici dans un contexte de recherche autour d'un système de *TAO personnelle*. L'utilisateur d'un tel système est un rédacteur unilingue qui souhaite pouvoir rédiger dans sa langue maternelle et être traduit pour un coût et dans un délai raisonnables. Pour la réalisation, on doit respecter trois contraintes très fortes. D'abord, la qualité de la traduction obtenue doit être suffisante pour ne pas devoir réviser (le rédacteur n'est pas un réviseur). Ensuite, on ne peut pas utiliser de base de connaissances pour obtenir automatiquement une compréhension explicite du document (il est impossible de connaître *a priori* le domaine sémantique de celui-ci). Enfin, on ne peut pas contraindre le rédacteur à utiliser un langage contrôlé, ni même un sous-langage déterminé (la rédaction doit être la plus libre possible).

Le respect des deux dernières contraintes ne permet pas la réalisation d'un système autonome de traduction avec lequel les analyses (syntaxique, logico-sémantique) du texte en langue source reflètent l'intention de l'auteur. Le système produira en effet, à certains stades du traitement, des représentations qui refléteront des ambiguïtés que les connaissances grammaticales « fortes » et les connaissances sémantiques « faibles » du système ne permettront pas de résoudre automatiquement. Or, si l'on veut obtenir une traduction de qualité, il faut obtenir une représentation désambiguïsée de ce que l'on veut traduire. La solution que nous envisageons au GETA est de réaliser, dans le cadre du projet LIDIA, la maquette d'un système de traduction qui utilise le dialogue avec le rédacteur pour obtenir une représentation désambiguïsée de ce que l'on veut

traduire. Dans un tel système, le dialogue occupe donc une place privilégiée puisque c'est lui qui va permettre d'obtenir une traduction de qualité.

Dans le présent article, nous décrirons comment, à partir d'un résultat d'analyse, nous percevons l'organisation des traitements qui nous permettront d'obtenir une représentation débarrassée de trois types d'ambiguïtés. Pour chacun de ces types, nous montrerons quel genre de dialogue peut proposer le système. Nous verrons alors que l'on peut adapter la forme du dialogue à l'utilisateur. Pour un utilisateur novice on peut envisager un dialogue très explicite. En effet, nous allons parfois surprendre le rédacteur en lui posant certaines questions. Il faut alors que l'explicitation du phénomène linguistique soit très précise et suffisamment simple pour être comprise sans avoir à faire appel à des connaissances linguistiques pointues. Ce type de dialogue doit aider le rédacteur à reconnaître les schémas ambigus pour éviter leur emploi ou bien être conscient qu'il devra répondre à des questions les concernant. Ensuite, avec une certaine habitude et une amélioration de ses compétences, il pourra choisir un mode de dialogue plus synthétique, qui rendra la tâche de désambiguïsation plus rapide.

Situation : l'analyseur de LIDIA

Les traitements interactifs dans le projet LIDIA se divisent en deux catégories : les traitements de standardisation et les traitements de clarification (Blanchon 1990, Boitet 1990). Lorsque les traitements de standardisation ont été effectués sur une unité de traduction, une première étape d'analyse produit une structure « concrète » qui est un arbre décoré « multiniveau » (Vauquois 1978). La structure est dite « concrète », car on y retrouve directement le texte (mot des feuilles). Elle est « multiniveau » car les nœuds portent des décorations complexes qui représentent divers niveaux d'interprétation : le niveau des classes syntaxiques et des classes syntagmatiques, le niveau des fonctions syntaxiques et enfin le niveau des relations logiques et sémantiques. Une première phase de désambiguïsation interactive s'applique à cette structure.

Pour préciser les choses, voici une brève description de ces divers niveaux d'interprétation qui permettra de mieux comprendre notre stratégie de désambiguïsation.

> Le niveau des classes syntaxiques (classes terminales telles que : « NOM COMMUN », « ADJECTIF », « PRÉPOSITION », etc.) et des classes syntagmatiques (classes non terminales telles que : « GROUPE NOMINAL », « PHRASE INFINITIVE », « PHRASE VERBALE », etc.) donne le parenthésage de la phrase en groupes syntagmatiques. Les feuilles de l'arbre portent les classes syntaxiques et les nœuds non terminaux portent les classes syntagmatiques.
>
> Le niveau des fonctions syntaxiques donne le rôle syntaxique de chaque nœud dans le groupe auquel il est attaché directement (par exemple : « GOUVERNEUR », « RÉGISSEUR », « SUJET », etc.). Avec ce niveau, on peut très simplement transformer l'arborescence en une arborescence de « dépendance syntaxique ».

Le niveau des relations logiques et sémantiques peut être considéré comme celui de la construction logique de la phrase, en mettant en évidence les places des arguments attachés aux unités lexicales prédicatives et donnant l'interprétation sémantique des compléments et circonstants (par exemple : « DÉTERMINATION », « CIRCONSTANCIEL INSTRUMENTAL », « PREMIER ARGUMENT D'UNE UNITÉ LEXICALE PRÉDICATIVE », etc.).

L'arbre d'analyse de la phrase: « L'élève calcule cette intégrale par la méthode des résidus » tel qu'il est classiquement calculé au GETA, est fourni pour clarification.

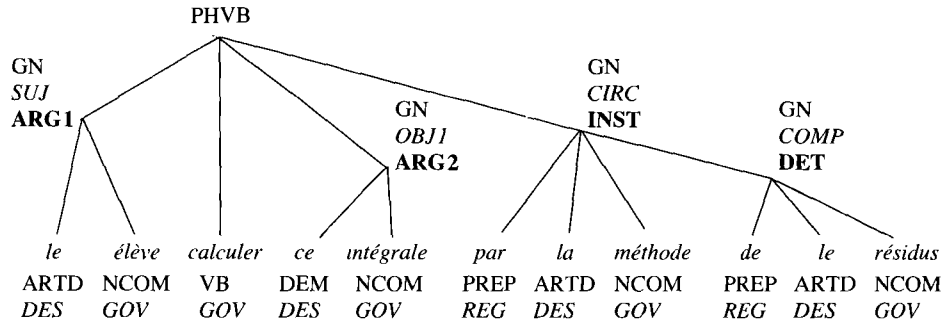


Figure 1 : Un arbre d'analyse multiniveau.

légende : NIVEAU 1, *NIVEAU 2*, **NIVEAU 3**

L'analyseur utilisé dans le cadre du projet LIDIA est un analyseur non déterministe « multiple » qui, pour une phrase donnée, fournit toutes les analyses vérifiant le modèle syntagmatique, syntaxique et logico-sémantique de l'analyseur. La figure 2 montre la forme d'une analyse produite pour une unité textuelle selon le modèle non déterministe. Pour la première phrase, il y a différentes analyses : le nœud père « UL(ULFRA),NUMFRA(1) » possède plusieurs fils UL(ULSOL) et, sous chaque nœud UL(ULSOL), on dispose d'un arbre d'analyse licite de la phrase.

On impose aussi, par rapport au modèle habituel, une contrainte supplémentaire à l'analyseur : le mot qui appartient à la décoration d'une feuille de l'arbre (*en italique dans la figure 1*) doit être une occurrence de la phrase analysée, et la projection du mot des feuilles de l'arbre selon cette composante mot de la décoration fournit la phrase analysée. Le respect de cette contrainte nous permet d'avoir un processus de génération des questions plus simple à mettre en œuvre et donc moins coûteux en logiciel et en temps. Le processus de génération du dialogue étant exécuté par la station de rédaction, on perçoit l'intérêt qu'il y a à minimiser l'importance des tâches à réaliser pour que leur exécution soit la plus transparente possible pour l'utilisateur. Il ne faut pas que

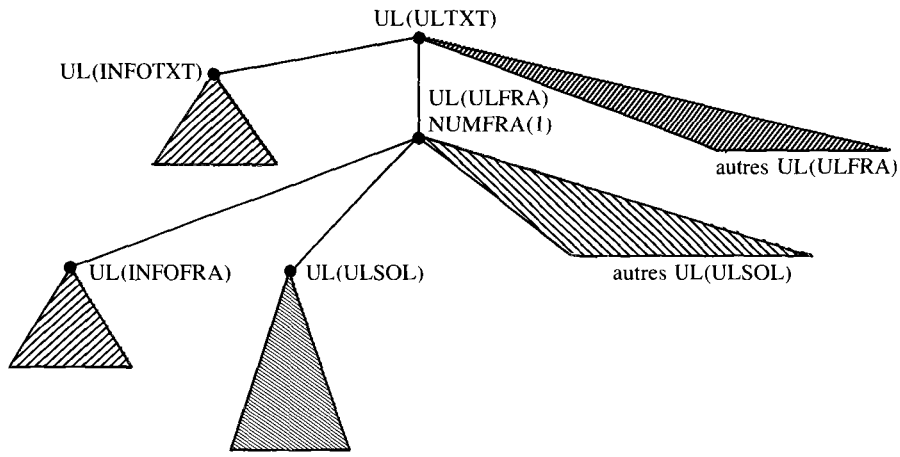


Figure 2 : Représentation d'une analyse non déterministe.

des tâches de fond trop importantes ralentissent le travail du rédacteur de façon significative.

Pour procéder à la désambiguïsation interactive en utilisant le résultat d'analyse comme support pour la préparation des questions, nous avons choisi d'utiliser trois propriétés qui permettent, pour une phrase, de différencier les divers arbres obtenus. Ces trois propriétés sont : l'étiquetage en classe syntaxique du mot des feuilles, l'étiquetage en fonction syntaxique (qui induit parfois des géométries différentes) et l'étiquetage en relation logique et sémantique des nœuds non terminaux de l'arbre. On va maintenant donner une définition de ces trois propriétés agrémentée d'exemples.

Propriétés choisies de la structure produite par l'analyseur

– La première propriété qui différencie les arbres d'analyse licite d'une phrase est l'étiquetage en classe syntaxique du mot des feuilles (désigné dans la suite par étiquetage en classe syntaxique) qui est, rappelons-le, le premier niveau d'interprétation de la structure profonde.

Le schéma suivant (figure 3) montre les différences entre les deux arbres produits par l'analyse de la phrase « Devant cette somme, il hésite » dans laquelle *devant* peut être interprété comme gérondif du verbe devoir ou bien comme préposition.

Comme on peut le voir, les arbres ont une géométrie différente (étiquetage en fonction syntaxique différent), mais ce qui nous intéresse ici c'est seulement la différence au niveau de l'étiquetage en classe syntaxique de l'occurrence *devant*.

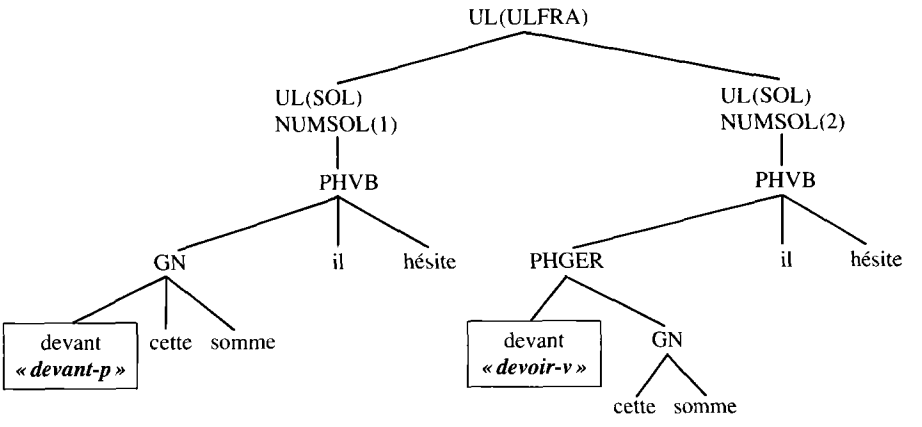


Figure 3 : Résultat d'analyse de la phrase « Devant cette somme, il hésite. » On observe un problème au niveau du choix de la classe syntaxique pour l'occurrence **devant**.

– La seconde propriété qui différencie les arbres d'analyse licite d'une phrase est la géométrie et l'étiquetage en fonction syntaxique des nœuds non terminaux (désigné par la suite par géométrie et étiquetage en fonction syntaxique). Les arbres ont maintenant tous le même étiquetage en classe syntaxique. Cette propriété correspond au deuxième niveau d'interprétation de la structure profonde.

Le schéma suivant (figure 4) montre les différences entre les deux arbres produits par l'analyse de la phrase « L'évolution de la structure du réseau et des investissements est étudiée » où l'on peut comprendre que l'on étudie l'évolution de la structure du réseau **et** l'évolution des investissements (NUMSOL(1)) ou bien que l'on étudie l'évolution de la structure du réseau **et** l'évolution de la structure des investissements (NUMSOL(2)).

– La troisième propriété qui différencie les arbres d'analyse licite d'une phrase est l'étiquetage en relation logique et relation sémantique des nœuds non terminaux (désigné par la suite par étiquetage en relation logico-sémantique). Elle correspond au troisième niveau d'interprétation de la structure profonde.

Le schéma suivant (figure 5) montre les différences entre les deux arbres produits par l'analyse de la phrase « Je vous parle de la tour Eiffel » où on peut comprendre que l'on nous parle depuis la tour Eiffel (NUMSOL(2)) ou bien que l'on nous parle au sujet de la tour Eiffel (NUMSOL(1)).

Après avoir décrit les propriétés qui nous intéressent, montrons comment, en exploitant leur présence dans le résultat d'analyse d'une phrase, nous pouvons résoudre certains types d'ambiguïtés.

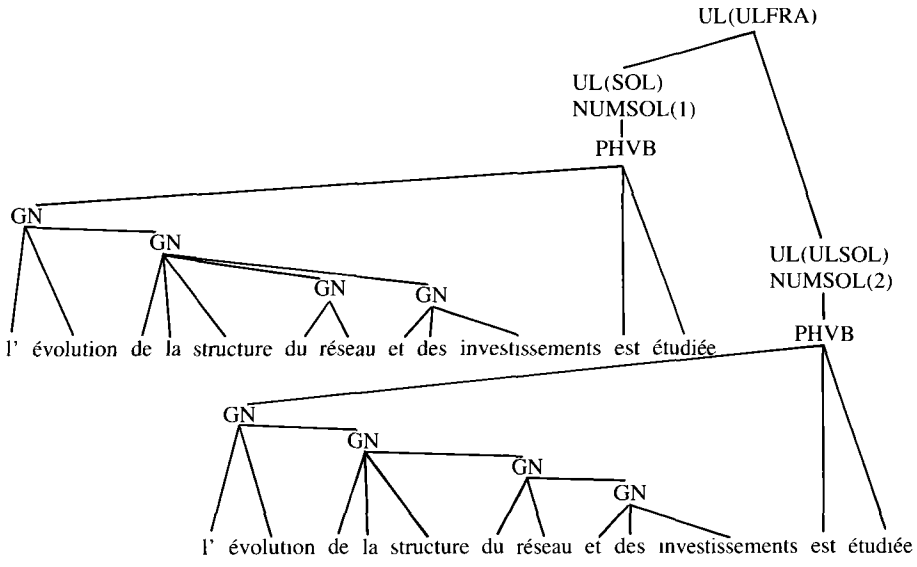


Figure 4 : Résultat d'analyse de la phrase « L'évolution de la structure du réseau et des investissements est étudiée. » On observe un problème au niveau du choix de la *géométrie* des arbres.

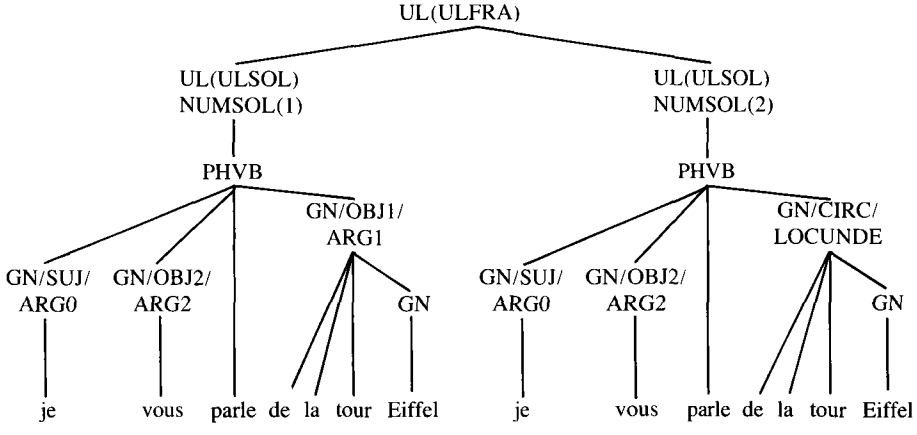


Figure 5 : Résultats d'analyse de la phrase « Je vous parle de la tour Eiffel. » On observe un problème au niveau du choix de la fonction du groupe nominal **de la tour Eiffel**.

Intérêt des propriétés choisies

Relation entre propriété et type d'ambiguïté

Si l'analyse indique un problème d'étiquetage en classe syntaxique, c'est l'ambiguïté de classe syntaxique que nous avons à résoudre. Cela signifie que l'analyseur, pour une occurrence donnée, n'a pas su déterminer quelle est sa classe syntaxique. Par exemple, dans la phrase « Le vétérinaire prend la patte de l'oiseau et **la** bague », on ne sait pas, *a priori*, si l'occurrence *bague* joue le rôle d'un verbe (**la** est alors pronom) ou d'un substantif (**la** est alors article).

Par contre, si l'analyse indique un problème de géométrie et d'étiquetage en fonction syntaxique, c'est l'ambiguïté de structure de constituants grammaticaux que nous devons résoudre. Cette ambiguïté de constituants grammaticaux est essentiellement liée à deux problèmes : la coordination et la subordination. La phrase « Il prend des classeurs et des cahiers noirs » présente un problème de géométrie pure, puisque l'on ne sait pas s'il faut comprendre que seuls les cahiers sont noirs ou bien que les classeurs sont noirs et les cahiers aussi. Pour la subordination, on peut citer la phrase : « Il est inscrit dans une école de cuisine lyonnaise » où on ne sait pas s'il s'agit d'une *école située à Lyon où on apprend la cuisine* ou bien d'une *école sise on ne sait où, où on apprend la cuisine de Lyon.*

Enfin, lorsque l'analyse indique un problème d'étiquetage en relation logico-sémantique, on a à résoudre deux types de problèmes : un problème de place des arguments attachés aux unités lexicales prédicatives (ex. : « Quel hôtel gère ce bureau? »), ou bien un problème d'interprétation sémantique des compléments et circonstants. Comme exemple de problème d'interprétation, on peut citer la phrase « Je vous parle de la tour Eiffel » où *de la tour Eiffel* peut être considéré comme circonstant de localisation (depuis la tour Eiffel), ou comme premier argument et premier objet du verbe parler (au sujet de la tour Eiffel).

Combinaisons possibles des propriétés

Bien sûr, une phrase peut donner lieu à plusieurs ambiguïtés, de même type ou de types différents. Tous les croisements sont effectivement possibles, comme le montrent les exemples suivants.

- Différents étiquetages en classes syntaxiques du mot des feuilles :
  - avec différentes géométries et étiquetages en fonctions syntaxiques et avec différents étiquetages en relations logico-sémantiques :

    Il nous parle de l'école de cuisine lyonnaise, prend le livre et le lit.

    On voudrait savoir comment interpréter
    « de », « de » : « depuis », « au sujet de »
    « école de cuisine lyonnaise » : « école lyonnaise », « cuisine lyonnaise »
    « le lit » : « pronom + verbe », « article + substantif »

- avec différentes géométries et étiquetages en fonctions syntaxiques et un seul étiquetage en relations logico-sémantiques :

  Il regarde le livre de l'école de cuisine lyonnaise et le lit.

- avec une seule géométrie et étiquetage en fonctions syntaxiques et avec différents étiquetages en relations logico-sémantiques :

  Il observe la photo de Paul et la classe.

  On voudrait savoir comment interpréter
  « de » : « prise par Paul », « représentant Paul », « appartenant à Paul »
  « la classe » : « pronom + verbe », « article + substantif »

- avec une seule géométrie et étiquetage en fonctions syntaxiques et un seul étiquetage en relations logico-sémantiques :

  Le vétérinaire prend la patte de l'oiseau et la bague.

  On voudrait savoir comment interpréter
  « la bague » : « pronom + verbe », « article + substantif »

- Différentes géométries et étiquetages en fonctions syntaxiques :
  - avec différents étiquetages en relations logico-sémantiques :

    Il parle de l'école de cuisine lyonnaise.

    On voudrait savoir comment interpréter
    « de » : « depuis », « au sujet de »
    « école de cuisine lyonnaise » : « école lyonnaise », « cuisine lyonnaise »

  - avec un seul étiquetage en relations logico-sémantiques :

    Il est inscrit dans une école de cuisine lyonnaise.

- Différents étiquetages en relations logico-sémantiques seulement :

  Quel hôtel gère ce bureau?

  On voudrait savoir comment interpréter les arguments du verbe
  « un hôtel gère-t-il le bureau? » ou « le bureau gère-t-il un hôtel? »

Définition d'un ordre d'utilisation des propriétés

Comme plusieurs propriétés peuvent apparaître dans un résultat d'analyse, nous allons exploiter, si elles se présentent, les propriétés dans l'ordre dans lequel nous les avons définies. C'est-à-dire l'étiquetage en classes syntaxiques, puis la géométrie et l'étiquetage en fonctions syntaxiques et enfin l'étiquetage en relations logico-sémantiques. La figure 6 donne une idée de la façon dont nous allons procéder. Le but du processus que nous allons décrire est la production d'une série de questions organisées hiérarchiquement qui permettent d'obtenir, en partant d'une liste d'analyses licites, l'analyse qui correspond aux intentions du rédacteur. Cet ordre d'exploitation des propriétés pourra, si nécessaire, être remis en question car, comme le montre la figure 6, l'organisation des traitements est fortement modulaire.

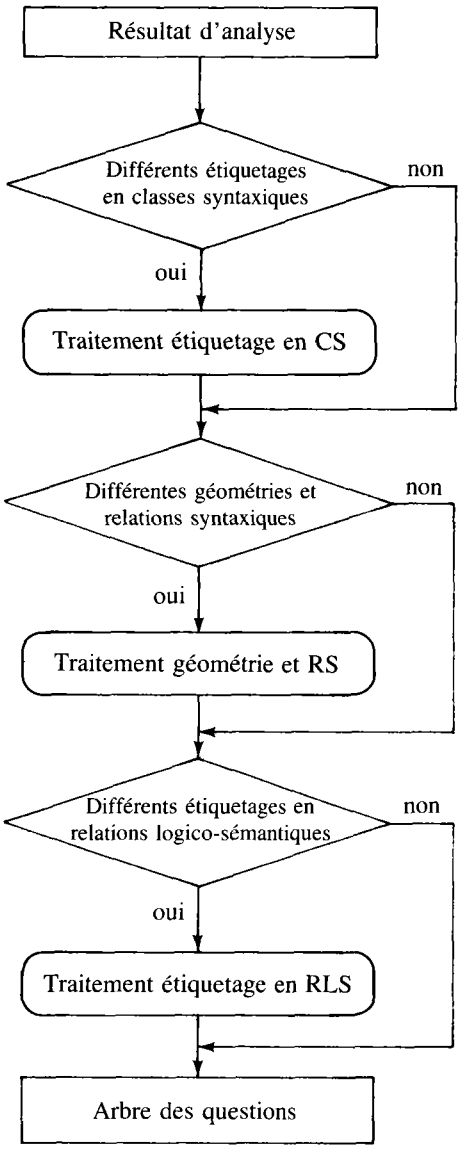


Figure 6 : Schéma du processus d'exploitation de l'arbre d'analyse.

On est libre d'examiner la présence de chacune des propriétés dans n'importe quel ordre. Mais il nous semble que l'organisation proposée a quelques fondements psychologiques et ergonomiques.

Pour le point de vue psychologique, lorsque l'on se place dans la situation de levée interactive d'ambiguïté, le rédacteur sera amené à réexaminer son texte sans se servir du contexte, comme le fait le système. Supposons par exemple qu'il soit occupé à rédiger un document qui parle du classement d'un certain nombre de clichés représentant des acteurs célèbres ou en train de le devenir et que dans un tel contexte il écrive la phrase « Il examine la photo de Pierre Latouche et la classe ». Dans ce texte, cette phrase ne pose aucun problème d'interprétation. Mais, sortie de tout contexte, elle est ambiguë, puisqu'on peut comprendre qu'il classe des photos ou qu'il examine une photo et une classe (un groupe d'élèves par exemple).

Hors contexte, la fabrication du sens (par un humain) semble procéder par restriction des classes syntaxiques des termes, puis des géométries et relations de coordination et de subordination qu'ils établissent ou non avec d'autres termes et enfin, par la qualité de ces relations. D'un point de vue psychologique, il semble donc convenable de procéder de manière analogue.

Le point de vue ergonomique renforce cette idée : il est préférable de réussir rapidement la désambiguïsation en posant les questions par ordre d'immédiateté de leurs réponses. C'est dire que la sélection de la bonne analyse doit procéder par « filtrages » successifs en utilisant d'abord des critères simples et des questions courtes, et ensuite seulement des critères plus fins et donc des questions plus complexes.

Ainsi, pour trouver la bonne interprétation de la phrase « Il atteint la grange et la ferme », il est plus naturel de demander si l'occurrence *ferme* doit être comprise comme le verbe *fermer* ou comme substantif que de demander si *et la ferme* doit être compris comme un groupe nominal coordonné au verbe *atteindre* ou bien comme une phrase verbale coordonnée au verbe *atteindre*. Il est donc normal de commencer par chercher le bon étiquetage en classes syntaxiques puis la bonne géométrie et le bon étiquetage en fonctions syntaxiques, et enfin le bon étiquetage en relations logico-sémantiques.

Utilisation des propriétés choisies en désambiguïsation

Généralités

Le système partitionnera donc l'ensemble des solutions en trois étapes successives, d'abord selon l'étiquetage en classes syntaxiques, puis selon la géométrie et l'étiquetage en relations syntaxiques et enfin, selon l'étiquetage en relations logico-sémantiques. On aboutira donc à un arbre des questions dont les niveaux successifs correspondront à cette organisation tripartite.

La figure 7 donne un exemple de découpage en classes et montre quel est l'arbre de décision que l'on peut construire pour hiérarchiser les questions.

Cette figure montre que les analyses A 1, A 2 et A 3 ont le même étiquetage en classes syntaxiques et que A 1 et A 2, tout en ayant le même étiquetage en fonctions syntaxiques, ont des étiquetages en relations logico-sémantiques différents; A 3, lui, a un autre étiquetage en fonctions syntaxiques. De même les analyses A 4 et A 5 ont le même étiquetage en classes syntaxiques, mais des étiquetages différents en relations syntaxiques.

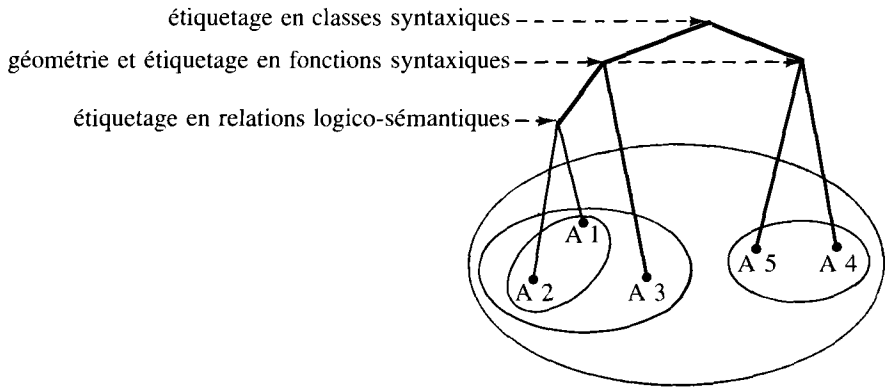


Figure 7 : Arbre de décision permettant de sélectionner la bonne analyse en fonction de la partition des analyses.

Exemple

Examinons la phrase « Le professeur de l'école de cuisine lyonnaise parle de la Tour d'Argent, prend le menu et le lit ». Cette phrase a seulement la prétention d'être un exemple démonstratif.

- Sans factorisation, l'analyse de cette phrase fournit huit solutions.
- On a schématiquement l'arbre de questions suivant :

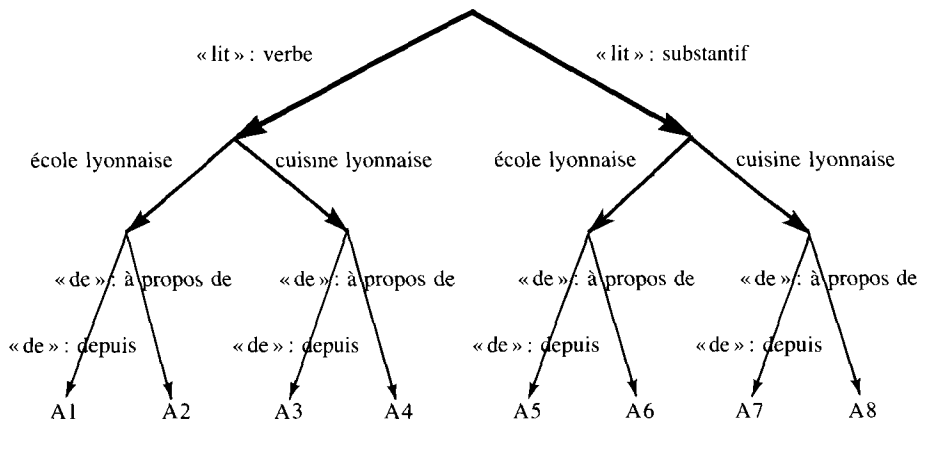


Figure 8

– Pour les questions on a les dialogues suivants :

– Pour la résolution des ambiguïtés de classes, on propose trois types de dialogue :

1) Une présentation grammaticale explicite du problème, pour se familiariser avec le système.

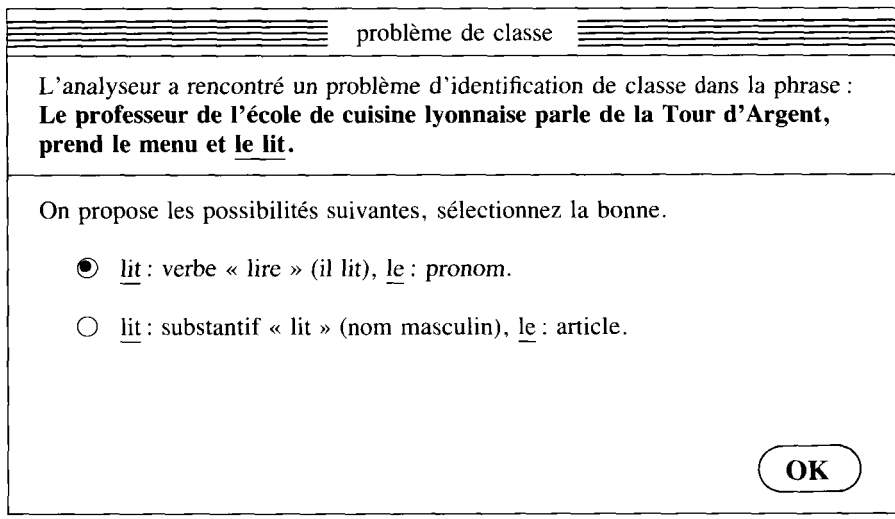

problème de classe
L'analyseur a rencontré un problème d'identification de classe dans la phrase :
Le professeur de l'école de cuisine lyonnaise parle de la Tour d'Argent, prend le menu et le lit.
On propose les possibilités suivantes, sélectionnez la bonne.
lit : verbe « lire » (il lit), le : pronom.
lit : substantif « lit » (nom masculin), le : article.
OK

Figure 9

2) Un paraphrasage de la partie de la phrase qui pose problème, l'explication grammaticale est implicite.

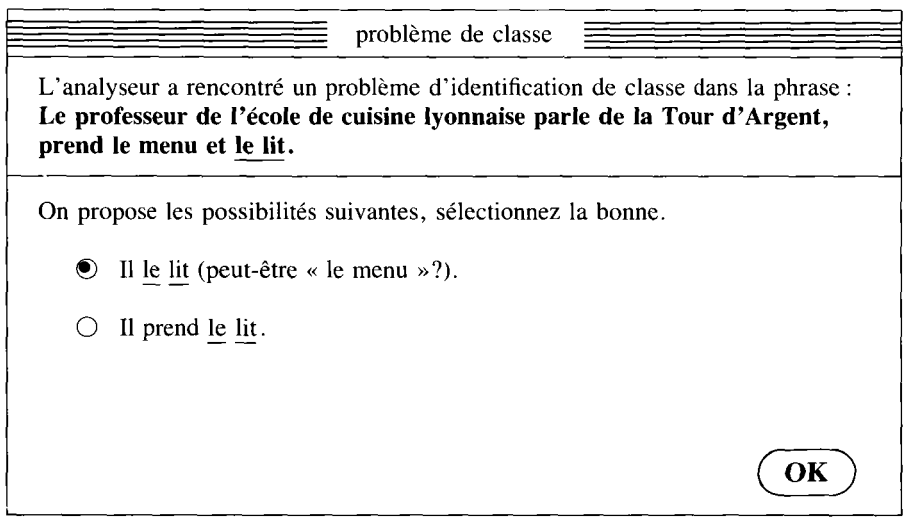

problème de classe
L'analyseur a rencontré un problème d'identification de classe dans la phrase :
Le professeur de l'école de cuisine lyonnaise parle de la Tour d'Argent, prend le menu et le lit.
On propose les possibilités suivantes, sélectionnez la bonne.
Il le lit (peut-être « le menu »?).
Il prend le lit.
OK

Figure 10

3) Une projection des classes syntaxiques de la partie de la phrase qui pose problème, destinée à un utilisateur qui a voulu se familiariser avec une présentation grammaticale des problèmes.

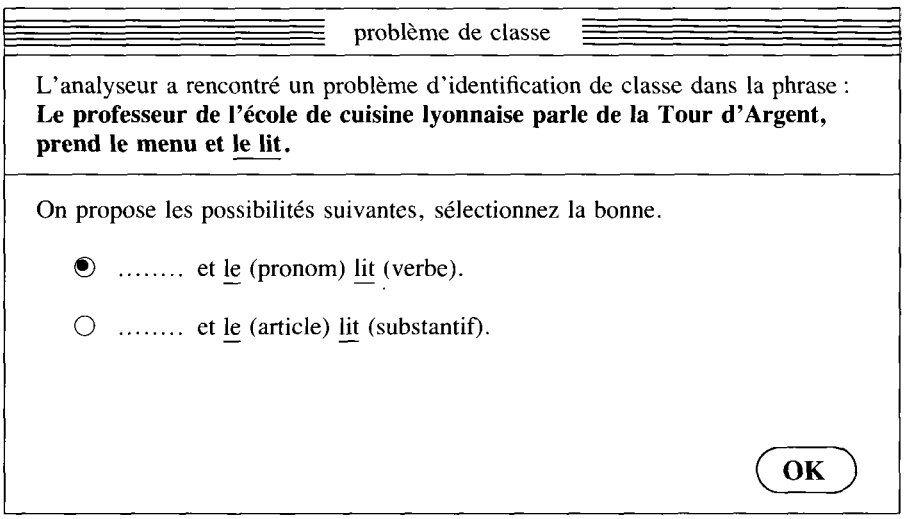

problème de classe
L'analyseur a rencontré un problème d'identification de classe dans la phrase :
Le professeur de l'école de cuisine lyonnaise parle de la Tour d'Argent, prend le menu et le lit.
On propose les possibilités suivantes, sélectionnez la bonne.
..... et le (pronom) lit (verbe).
..... et le (article) lit (substantif).
OK

Figure 11

– Pour la résolution des ambiguïtés de constituants grammaticaux, on peut proposer des présentations de type paraphrastique qui tiennent plus ou moins compte du contexte

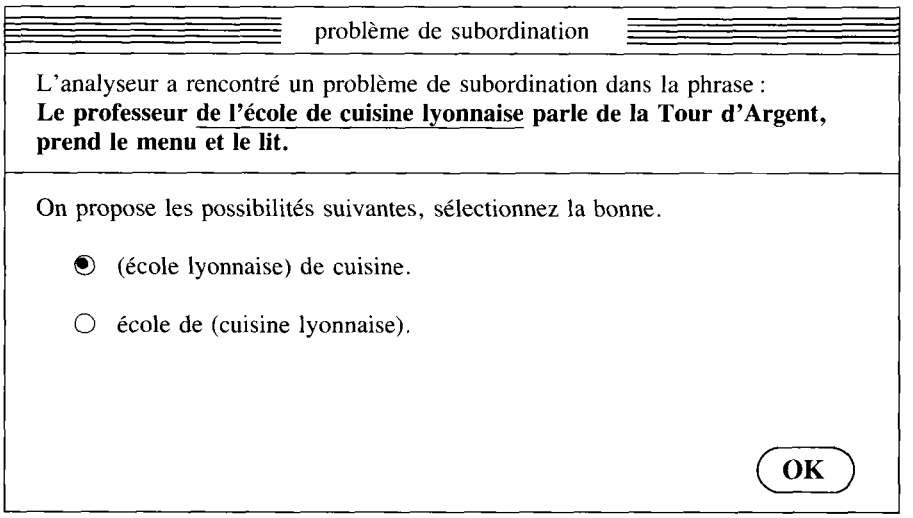

problème de subordination
L'analyseur a rencontré un problème de subordination dans la phrase :
Le professeur de l'école de cuisine lyonnaise parle de la Tour d'Argent, prend le menu et le lit.
On propose les possibilités suivantes, sélectionnez la bonne.
(école lyonnaise) de cuisine.
école de (cuisine lyonnaise).
OK

Figure 12

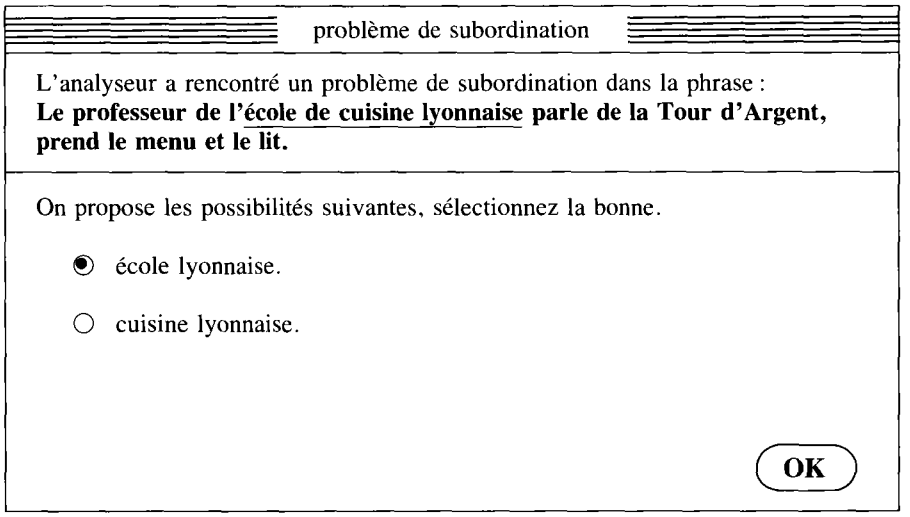

problème de subordination
L'analyseur a rencontré un problème de subordination dans la phrase :
Le professeur de l'école de cuisine lyonnaise parle de la Tour d'Argent, prend le menu et le lit.
On propose les possibilités suivantes, sélectionnez la bonne.
école lyonnaise.
cuisine lyonnaise.
OK

Figure 13

ou une présentation plus compacte qui exige peut-être un peu plus d'expérience du système.

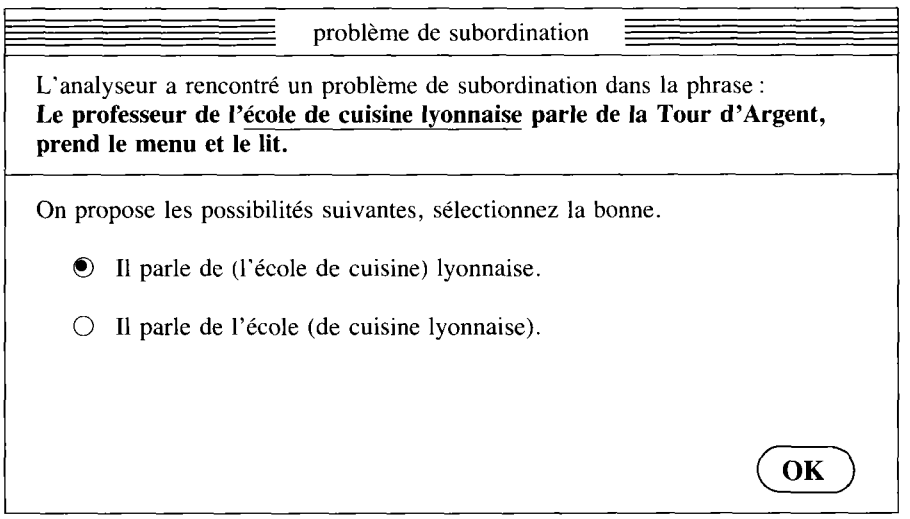

problème de subordination
L'analyseur a rencontré un problème de subordination dans la phrase :
Le professeur de l'école de cuisine lyonnaise parle de la Tour d'Argent, prend le menu et le lit.
On propose les possibilités suivantes, sélectionnez la bonne.
Il parle de (l'école de cuisine) lyonnaise.
Il parle de l'école (de cuisine lyonnaise).
OK

Figure 14

- Pour la résolution des ambiguïtés de relation logico-sémantique, on peut ici proposer des niveaux très différents de dialogue :
  1) Un niveau où on remplace l'occurrence qui permet de mettre en évidence la place des arguments ou d'introduire les compléments ou les circonstants. Ici,

il s'agit de remplacer *de* qui permet d'introduire, soit l'argument 1 du verbe parler soit un circonstant de localisation.

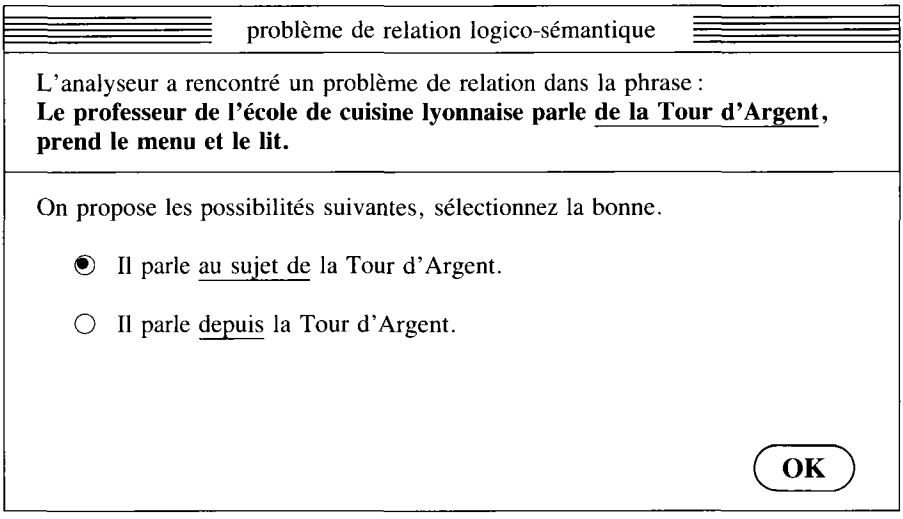

problème de relation logico-sémantique
L'analyseur a rencontré un problème de relation dans la phrase :
Le professeur de l'école de cuisine lyonnaise parle de la Tour d'Argent, prend le menu et le lit.
On propose les possibilités suivantes, sélectionnez la bonne.
Il parle au sujet de la Tour d'Argent.
Il parle depuis la Tour d'Argent.
OK

Figure 15

2) Un niveau qui précise en plus l'étiquetage de l'occurrence *de* qui permet d'apprendre les représentations du système.

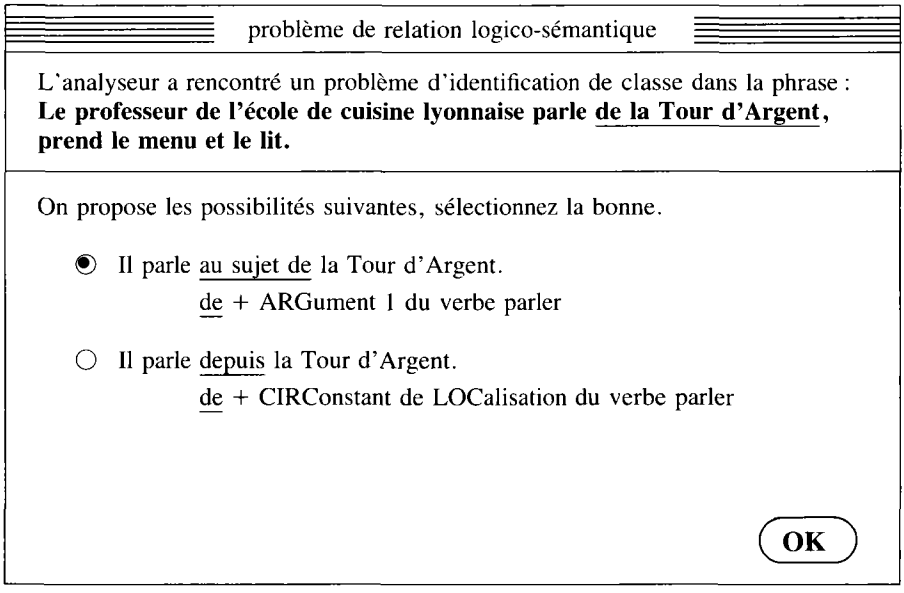

problème de relation logico-sémantique
L'analyseur a rencontré un problème d'identification de classe dans la phrase :
Le professeur de l'école de cuisine lyonnaise parle de la Tour d'Argent, prend le menu et le lit.
On propose les possibilités suivantes, sélectionnez la bonne.
Il parle au sujet de la Tour d'Argent.
de + ARGument 1 du verbe parler
Il parle depuis la Tour d'Argent.
de + CIRConstant de LOCalisation du verbe parler
OK

Figure 16

3) Un niveau pour un utilisateur spécialiste.

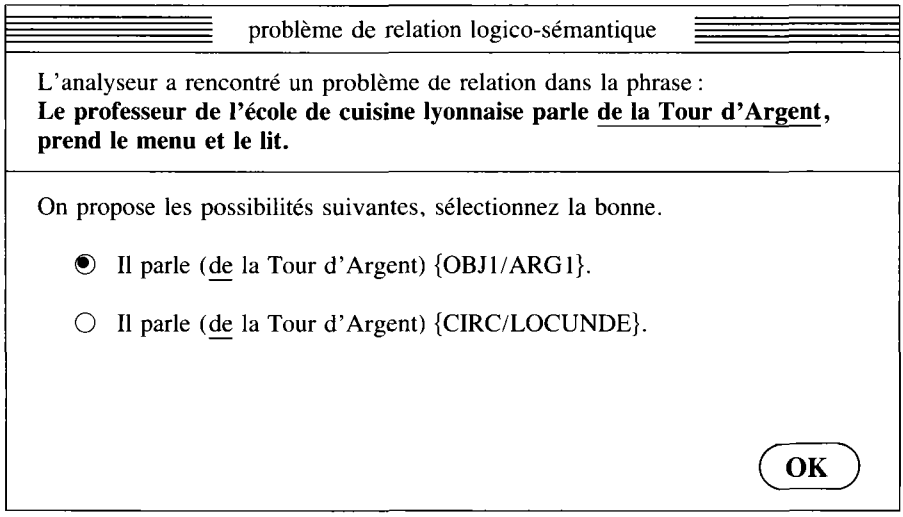

problème de relation logico-sémantique
L'analyseur a rencontré un problème de relation dans la phrase :
Le professeur de l'école de cuisine lyonnaise parle de la Tour d'Argent, prend le menu et le lit.
On propose les possibilités suivantes, sélectionnez la bonne.
Il parle (de la Tour d'Argent) {OBJ1/ARG1}.
Il parle (de la Tour d'Argent) {CIRC/LOCUNDE}.
OK

Figure 17

Conclusion

Nous avons défini trois propriétés de la structure arborescente concrète, mais profonde (en fait, multiniveau) produite par un analyseur non déterministe (dans le formalisme du GETA). Nous avons vu comment exploiter ces propriétés dans le cadre d'un système de « TAO personnelle » pour produire un dialogue de désambiguïsation qui soit réglable en fonction des compétences de l'utilisateur. Les ambiguïtés que l'on résout sont, on le rappelle, les ambiguïtés de classe syntaxique, les ambiguïtés de coordination et de subordination, et enfin les ambiguïtés de relation logique et sémantique. Mais, pour obtenir une représentation totalement désambiguïsée il restera d'autres problèmes à régler.

C'est ainsi qu'il faut encore résoudre les problèmes de polysémie (pure et induite par la langue cible), les problèmes de référence pronominale ou adjectivale et les problèmes d'ellipse, soit pure (toutes les fois qu'un terme de phrase ou de groupe n'est pas repris, souvent dans la coordination ou la comparaison), soit dans les termes composés (lorsque l'on ne reprend pas complètement un terme composé pour abréger ou alléger le texte, le terme est remplacé par son gouverneur).

Les phases de désambiguïsation qui traitent de ces problèmes sont réalisées soit en amont soit en aval des traitements que nous avons décrits ici. Pour information, on peut dire que pour les polysémies, la notation du problème sera inscrite dans l'arbre dans l'étiquette de sens associée à chaque occurrence, et les informations nécessaires à la désambiguïsation seront fournies par la base lexicale. Pour les problèmes de référence pronominale, il faudra conserver une liste de toutes les occurrences qui

peuvent être reprises en associant à chacune d'elles des informations grammaticales (genre, nombre...) et des informations lexicales provenant du dictionnaire (les rections syntaxiques et traits sémantiques). Ces informations permettront de retrouver les occurrences pertinentes candidates à la reprise. Quant au problème de l'ellipse dans les termes composés, une liste des termes composés déjà rencontrés ou des informations à propos du domaine du document permettront de trouver une liste des candidats possibles.

Bibliographie

BLANCHON H. (1990b) « HyperAriane : architecture de la partie *micro* d'un système de TAO personnelle », *Mémoire de DEA*, juin 1990, 63 p.

BLANCHON H. (1990a) « LIDIA-1 : Un prototype de TAO personnelle pour rédacteur unilingue », *Proc. Avignon-90, conférence spécialisée Le Traitement des langues naturelles et ses applications*, 28 mai–1er juin, 1990, p. 51-60.

BOITET Ch. (1990b) « Vers la TAO personnelle : le projet LIDIA du GETA », à paraître dans *la tribune sur les industries de la langue*, 12 p.

BOITET Ch. (1990a) « Towards Personal MT : On Some Aspects of the LIDIA Project », *Proc. COLING-90*, Helsinki, 20-25 août 1990, 10 p.

BOITET Ch. (1989b) « Speech Synthesis and Dialog Based Machine Translation », *Proc. ATR Symp. on Basic Research for Telephone Interpretation*, Kyoto, Dec. 1989, 6-5-1–6-5-22.

BOITET Ch. (1989a) « Motivation and Architecture of the LIDIA Project », *Proc. MTS-89*, Munich, 16-18 août 1989, 5 p.

BOITET Ch. et BLANC E. (éds.) (1990) *DBMT-90 : Post-COLING Seminar on Dialogue-Based MT (Machine Translation of/with Dialogues)*, Le Sappey (France), 26-28 août 1990, 328 p.

BROWN R. D. et NIREMBURG S. (1990) « Human-Computer Interaction for Semantic Disambiguation », *Proc. COLING-90*, Helsinki, 20-25 août 1990, 6 p.

HUTCHINS W. J. (1989) « Out of the Shadows, A Retrospect of Machine Translation in the Eighties », Presented at *Computer and Translation 89*, Tbilissi, Georgia, novembre-décembre 1989, 13 p.

LEPAGE Y. (1988) « Ambiguities and Second Generation MT Systems », *First European Conference on Information Technology for Organisational Systems*, Athens, 16-20 mai 1988, 6 p.

MARUYAMA H. et WATANABE H., OGINO S., (1989) « An Interactive Japanese Parser for Machine Translation », *Proc. COLING-90*, Helsinki, 2-25 août 1990, 7 p.

PECKHAM J. (ed) (1989) « Recent Developments and Applications of Natural Language Processing », Unicom Seminars Lilited, dans la collection *Applied Information Technology Reports*, 300 p.

TOMITA M. (1985) *Feasibility Study of Personal/Interactive Machine Translation System*, Center for Machine Translation, Carnegie Mellon University, mai 1985, 13 p.

TOMITA M. (1984) « Disambiguating Grammatically Ambiguous Sentences by Asking », *Proc. COLING-84*, p. 476-480.

VAUQUOIS B. (1978) « Description de la structure intermédiaire », Communication présentée au *colloque de Luxembourg*, 17 et 18 avril 1978, 25 p.

WHITELOCK P. J., WOOD M. M., CHANDLER B. J., HOLDEN N., HORSFALL H. J. (1986). « Strategies for Interactive Machine Translation : the Experience and Implications of the UMIST Japanese Project », *Proc. COLING-86*, p. 329-334.

ZAJAC R. (1986) *Étude des possibilités d'interaction homme-machine dans un processus de traduction automatique*, Thèse, INPG, Grenoble, juillet 1986, 259 p.

ZAJAC R. (1988) « Interactive Translation : a New Approach », *Report GETA & ATR*, Osaka, mai 1988, and *Proc. COLING-88*, Budapest, août 1988.

4

Dictionnaires électroniques dans un environnement traductionnel avancé

Ulrich HEID

Université de Stuttgart, Stuttgart, Allemagne

Avec la croissance du marché de la documentation, les outils informatiques deviennent de plus en plus importants; d'autre part, il subsiste, paraît-il, un écart considérable entre la performance des outils existants et les exigences des utilisateurs potentiels. Devant une telle situation, l'investissement dans les outils devrait s'accompagner de réflexions sur la compatibilité des systèmes avancés d'aujourd'hui avec les exigences du domaine, telles qu'elles se dessinent à travers les projets de recherche en cours.

Le domaine des ressources lexicales et terminologiques a déjà sérieusement entamé la discussion sur les questions de compatibilité et de réutilisation. Dans ce domaine, il existe déjà un certain nombre d'outils, pour la plupart interactifs, dont les chances d'intégration avec des applications avancées, telles que la génération automatique de textes, la traduction automatique, etc., soulèvent des doutes sérieux. Il faut alors s'interroger sur les caractéristiques que des ressources lexicales plus flexibles devraient avoir.

L'exposé part de ce scénario et donne une description rapide d'une base de données lexicale et terminologique conçue récemment en coopération avec l'entreprise Krupp Industrietechnik (Duisburg) qui constitue l'élément central d'une station de travail intégrée. À partir de travaux faits dans Polygloss, un projet de génération multilingue et d'EUROTRA-7, une étude de faisabilité sur la réutilisation de ressources lexicales, il faudra montrer les conditions nécessaires à une intégration d'outils interactifs et automatiques. L'accent sera mis sur les descriptions linguistiques nécessaires à une telle flexibilité.

5

Le module lexical dans un système de traduction automatique*

Liesbeth DEGAND, Guy EVERAERT, Jacques LEROT

Projet GENESE, Institut de linguistique, Louvain-la-Neuve, Belgique

Introduction

La génération et la traduction automatique de textes sont des procédés complexes dont la réalisation implique une architecture modulaire (modules syntaxique, discursif, sémantique et lexical). Ces modules sont en interaction permanente.

Dans le présent article, nous nous proposons de présenter le module lexical tel que nous l'envisageons dans notre système de génération automatique de textes.

Afin d'appréhender le lexique dans toute sa complexité, nous posons à l'intérieur du module lexical l'existence de trois niveaux :

- le modèle du monde (*world modelling structure*),
- le lexique conceptuel (*concept lexicon*),
- le lexique spécifique à chaque langue particulière (*language specific lexicon*).

Nous allons voir comment ces trois niveaux agissent les uns sur les autres pour fournir les lexèmes adéquats dans le processus de la traduction.

* La présente recherche, financée par la Communauté française de Belgique sous la convention n° 89-94/137, se fait dans le cadre d'un projet consacré à la génération de phrases françaises et néerlandaises à partir de représentations sémantiques. Il s'inscrit dans les efforts internationaux visant à développer un système de traduction multilingue (projet GENESE).

Le modèle du monde

Toute assertion fait appel aux connaissances générales appelées aussi connaissances du monde. Le *modèle du monde* est la composante qui représente ces connaissances générales[1].

Cette composante est représentée sous forme d'une hiérarchie où les éléments sont reliés entre eux de façon superordonnée et sousordonnée (relation qui sont exprimées par des arcs *une-sorte-de* (AKO — a kind of —)[2]. Le schéma suivant représente une telle organisation.

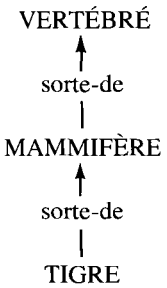


Figure 1

Cette représentation ne fournit qu'une information partielle à propos du monde. On décrira davantage les nœuds de la hiérarchie (appelés *types*) en spécifiant leurs propriétés typiques. On associe ainsi à chaque type un ensemble de paires attribut-valeur appelés *traits typés* (*feature types*) (Aït-Kaci 1986), (Emele, Zajac 1990). On représentera une telle paire de la manière suivante :

TYPE [attribut : valeur]

1. En réalité nous n'avons jamais besoin de toutes les connaissances du monde pour interpréter un fait, mais d'un sous-ensemble précis des connaissances générales (un sous-monde). Le modèle du monde est donc en fait un modèle du sous-monde (Nirenburg, Raskin 1987).
2. Pour une description plus complète des arcs, *cf.* Thayse 1988, chap.3.

Par exemple, le type « TIGRE » sera décrit à l'aide d'un ensemble de paires attribut-valeur tels que [possède : griffes], [est-une : espèce-en-danger], [sorte-de : mammifère], etc. On obtient le schéma suivant :

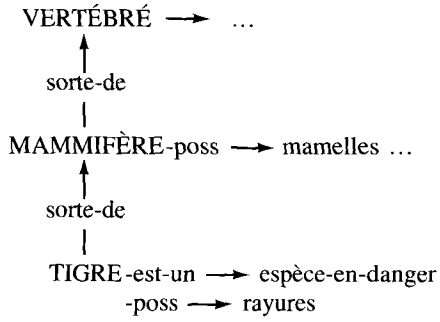


Figure 2

Ce système définit des types capables de représenter n'importe quelle réalité. Il permet donc d'intégrer des éléments mythiques imaginaires tels que, par exemple, les licornes, les griffons, etc.

Il y a plusieurs façons d'envisager ces hiérarchies qui varient selon les auteurs. Par exemple, pour Nirenburg et Carlson (1990), la construction se subdivise en deux sous-hiérarchies : ÉVÉNEMENT et OBJET.

Le type OBJET introduit les concepts OBJET-PHYSIQUE, OBJET-MENTAL et OBJET-SOCIAL. Les objets physiques peuvent à leur tour être classés en ENTITÉ SÉPARABLE, MATÉRIEL et LIEU. Parmi les objets mentaux, on reconnaît les OBJETs-ABSTRAITs et les OBJETs-REPRÉSENTATIFs (comprenant les OBJETs-MATHÉMATIQUEs, les OBJETs-LANGAGIERs, les ICÔNEs et les OBJETs-PICTURAUX). Finalement, les objets sociaux sont des ENTITÉs-GÉOPOLITIQUEs et des ORGANISATIONs.

Le type ÉVÉNEMENT est également subdivisible en un monde physique, mental et social. Les ÉVÉNEMENTs-PHYSIQUEs contiennent le CHANGEMENT-DE-LIEU et l'APPLICATION-DE-FORCE, les événements mentaux sont subdivisés en ÉVÉNEMENTs-COGNITIFs et ÉVÉNEMENTs-PERCEPTUELs, le monde social contient quant à lui des ÉVÉNEMENTs-COMMUNICATIFs.

Le sommet de la hiérarchie est représenté de la façon suivante (d'après Carlson L., Nirenburg S. 1990) :

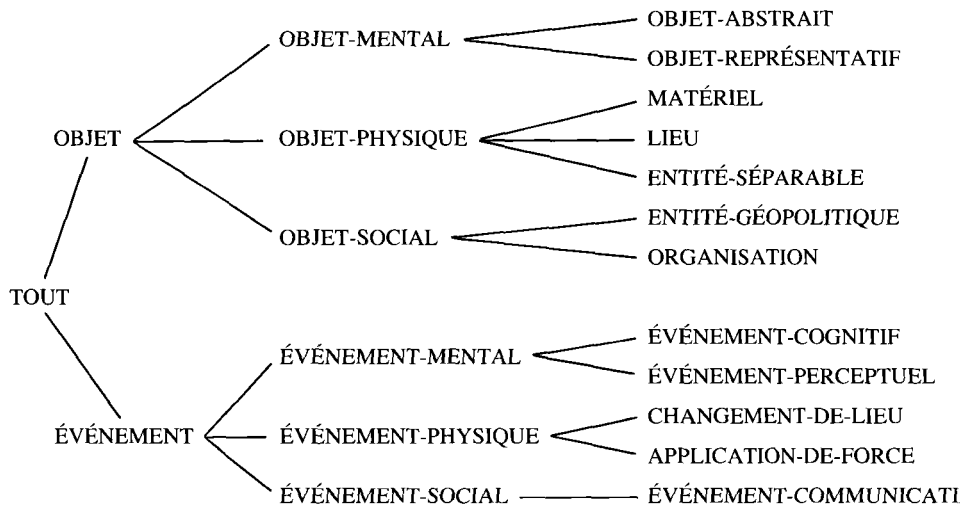


Figure 3

En résumé, nous pouvons dire que le modèle du monde est en quelque sorte un réservoir d'informations pertinentes qui sont à la base du processus de conceptualisation que nous décrirons dans la section suivante.

Il est important de noter que le modèle du monde est dynamique et qu'il peut être modifié et adapté en fonction des besoins de l'utilisateur. La plupart des changements auront lieu dans la partie inférieure de la structure hiérarchique qui contient les données relevant des domaines particuliers.

Le lexique conceptuel

On peut définir le lexique conceptuel comme l'organisation des concepts qui se trouvent à la base de la sémantique du langage naturel (d'après Pustejovsky 1990 : 2).

Les concepts sont le reflet du modèle du monde et sont soit de simples projections d'un type, soit le résultat d'une combinaison de différents types. Par exemple, le concept de « brique » est en fait la conceptualisation de la conjonction des types « artefact » et « construire ».

Le concept est représenté au moyen d'un graphe canonique, c'est-à-dire un réseau sémantique organisé autour d'une tête (Sowa 1984).

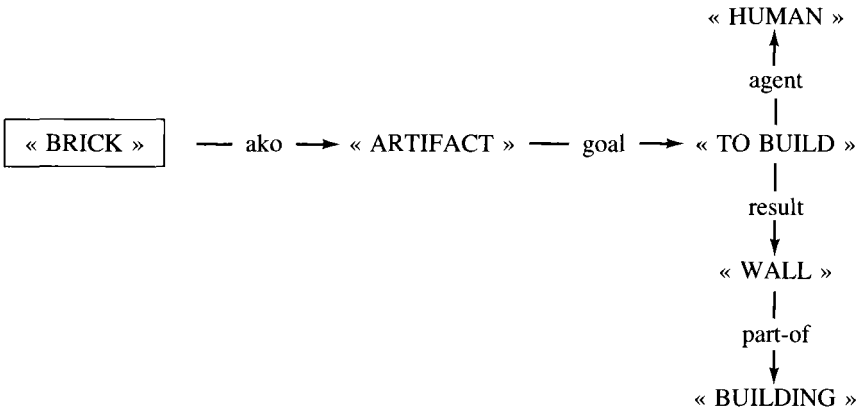


Figure 4

Ce réseau sémantique représente un concept, c'est-à-dire une unité de signification qui peut être définie au moyen d'une paraphrase, que nous appellerons *paraphrase définitoire*. Les arcs entre les nœuds ont une direction définie et sont étiquetés par des relations qui font partie de la définition du concept. Les relations indiquent comment deux nœuds sont reliés. Le concept sera réalisé par une unité lexicale lors du processus de lexicalisation (*cf. infra*).

Dans la traduction, le lexique conceptuel fonctionne comme une interlangue. Il est le pivot dans la recherche de la bonne traduction.

Le lexique conceptuel a donc pour but de rendre l'information sémantique explicite et accessible sous forme de réseaux sémantiques représentant les paraphrases qui définissent les concepts.

La paraphrase définitoire n'épuise pas la définition du concept. En effet, le concept est également associé à un certain savoir ajouté (*additional knowledge*) qui reste constant à l'intérieur d'une même culture. Les connotations et les métaphores reposent principalement sur ces connaissances ajoutées.

Le concept comporte deux parties :

a) la paraphrase définitoire, dont les éléments de sens sont hérités du modèle du monde comprenant les types et leurs relations;
b) le savoir ajouté.

Ces données sont reprises dans le schéma simplifié suivant :

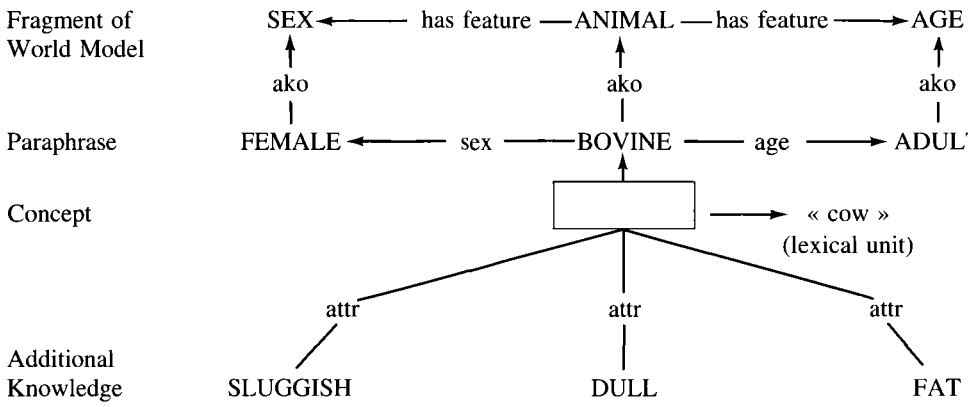


Figure 5

Notons que les concepts qui héritent d'un grand nombre de propriétés communes sont placés à un même niveau dans l'arbre hiérarchique, on peut en quelque sorte saisir le concept dans son *champ sémantique*.

Le lexique spécifique

Le lexique conceptuel met un concept en relation avec sa signification; il diffère du lexique particulier à chaque langue, qui lui met les concepts en relation avec les unités du vocabulaire. Le lexique conceptuel est donc une interlangue neutre par rapport à toute langue naturelle, mais il se trouve à la base du lexique spécifique de chaque langue. Le passage du lexique conceptuel au lexique spécifique est réalisé grâce au processus de lexicalisation qui fait correspondre chaque concept à un lexème défini.

Si un concept n'est pas lexicalisé dans une langue donnée, on peut néanmoins l'exprimer à l'aide d'une paraphrase. Cette paraphrase exprime en fait le premier concept lexicalisé supérieur dans la hiérarchie, auquel on ajoute les traits typiques manquants du concept non lexicalisé.

Par exemple, le lexème *embouteiller* est la lexicalisation de la construction conceptuelle formée des primitifs « mettre dans » et « bouteilles ». Ce lexème n'existe pas en allemand, il n'y a pas un verbe tel que « einflaschen ». On prend alors le concept lexicalisé supérieur et on y ajoute les traits typiques manquants de *embouteiller*.

En allemand, nous obtenons alors *in Flaschen* (trait typique) *einfüllen* (concept complexe supérieur). Le graphe conceptuel de *embouteiller* ressemblerait au schéma (simplifié) représenté ci-dessous.

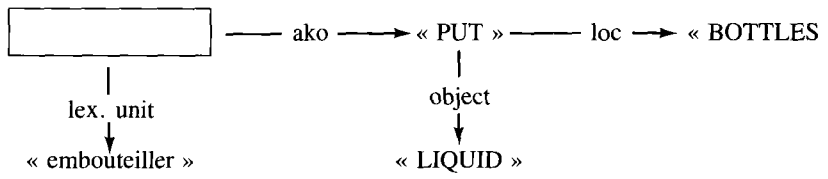


Figure 6

Les choses sont un peu plus compliquées lorsqu'on traduit dans le sens inverse, c'est-à-dire d'une langue où le concept n'est pas lexicalisé (*in Flaschen einfüllen* en allemand) vers une langue où le concept est lexicalisé mais où la paraphrase existe également (*embouteiller* ou *mettre en bouteilles/verser des bouteilles* en français). Lorsqu'il s'agit de traduction de textes écrits, nous prenons l'option de toujours choisir le terme le plus spécifique[3].

Un autre problème risque de se poser lorsqu'on veut modifier la paraphrase. Prenons le cas du lexème néerlandais *schimmel* paraphrasé en français par *étalon blanc*. Que se passe-t-il lorsqu'on veut traduire en néerlandais *un étalon blanc*? Dans ce cas, la modification porte en fait sur une partie seulement de la paraphrase, on dit de la couleur du cheval qu'elle est très blanche. Pour rendre cette notion de modification partielle en néerlandais, il faut faire éclater le concept et ne modifier que la partie concernée du concept, dans notre cas la couleur, ce qui donne *een heel witte schimmel*. Il ne s'agit pas ici d'un pléonasme puisqu'un *schimmel* n'est pas toujours **très** blanc. La modulation du modificateur et la décomposition du concept en sa paraphrase nous offrent donc une solution élégante à ce problème.

Conclusion

Nous avons émis l'hypothèse que le lexique conceptuel était un niveau clé pour la traduction automatique. Nous avons proposé de représenter ces connaissances sous forme de réseaux sémantiques, qui doivent permettre toutes les applications informatiques et le développement de logiciels d'aide à la traduction. À l'intérieur de ce formalisme, nous avons proposé des mécanismes pour choisir une traduction adéquate.

3. Cette même approche est défendue par Nogier (1991).

Bibliographie

AIT-KACI H. (1986) « An Algebraic Semantis Approach to the Effective Resolution of Type Equation », *Theoretical Computer Science 45*, p. 193-351.

BOGURAEV B. and J. PUSTEJOVSKY (1990) « Lexical Ambiguity and the Role of Knowledge Representation in Lexicon Design », *Proceedings of Coling '90*.

CARLSON L. and S. NIRENBURG (1990) *World Modeling for NLP*, Technical report CMU-CMT-90-121.

COSTERMANS J. (1980) *Psychologie du langage*, Bruxelles, Pierre Mardaga.

EMELE M. and ZAJAC R. (1990) *A Fixed Point Semantics for Type Systems*, Stuttgart, ms Polygloss.

EVENS M.W. (ed.) (1988) *Relational Models of the Lexicon, Representing Knowledge in Semantic Networks*, Cambridge University Press.

LEROT J. (1991) *Outline of the GENESE Model of Semantics Based Text Generation System*, Louvain-la-Neuve, ms Projet GENESE.

MEYER I., ONYSHKEVYCH B. and L. CARLSON (1990) *Lexicographic Principles and Design for Knowledge-Based Machine Translation*, Technical Report CMU-CMT-90-118.

NIRENBURG S. and V. RASKIN (1987) « The Subworld Concept Lexicon and the Lexicon Management System », *Computational Linguistics*, 13, 3-4, p. 276-289.

NOGIER J-F. (1991) *Génération automatique de langage et graphes conceptuels*, Paris, Hermès.

PUSTEJOVSKY (1989) *The Generative Lexicon*, ms. Brandeis University.

SOWA J.F. (1984) *Conceptual Structures*, Reading, Massachussets, Addison-Wesley Publishing Company.

SOWA J.F. (1988) « Using a Lexicon of Canonical Graph in a Semantic interpreter », in EVENS M.W. (ed.), (1988), p. 113-138.

THAYSE A. *et al.* (1988) *Approche logique de l'Intelligence Artificielle*, vol. 1, Paris, Dunod.

WIERZBICKA A. (1985) *Lexicography and Conceptual Analysis*, Ann Arbor, Karoma Publishers.

6

Pour une approche interactive au problème de la traduction automatique

Eric WEHRLI*

Département de linguistique générale et de linguistique française, Université de Genève, Genève, Suisse

Introduction

Au cours des quelques dernières années, on a pu observer un intérêt croissant pour des méthodes interactives dans le domaine de l'analyse du langage en général et dans celui de la traduction automatique en particulier. Contrairement aux systèmes non interactifs, ou autonomes, qui ne permettent aucune intervention de leur utilisateur en cours de processus, un système interactif est basé sur l'idée d'une collaboration entre un programme et son utilisateur. Autrement dit, ce type d'approche cherche à associer génie logiciel et génie humain plutôt que de les mettre en compétition. Concrètement, un système interactif tel que nous l'entendons ici est un système qui, lorsqu'il fait face à un problème qu'il est incapable de résoudre par lui-même, se réserve la possibilité de consulter son utilisateur pour, par exemple, lui demander de choisir entre différentes solutions ou lui permettre de corriger une éventuelle erreur. Les systèmes interactifs sont donc basés sur l'idée d'une collaboration directe, en temps réel, entre le progamme de traduction et son utilisateur, ce dernier conservant par conséquent un certain contrôle sur le déroulement du processus de traduction.

* Je tiens à remercier Robin Clark et Catherine Walther pour leurs commentaires et suggestions. La recherche décrite dans cet article est supportée en partie par un subside du Fonds national suisse de la recherche scientifique (subside no 11-25362.88).

Dans cet exposé, nous présenterons quelques-uns des arguments en faveur de systèmes interactifs dans le domaine de la traduction, et montrerons en quoi de tels systèmes s'opposent aux systèmes traditionnels de traduction automatique, non seulement du point de vue de l'utilisateur, mais également du point de vue plus technique qui est celui du concepteur. Nous aborderons ensuite certains problèmes liés au développement de systèmes interactifs et chercherons en particulier à répondre à des questions telles que :

- Quels types de problèmes devraient être traités de façon interactive?
- À quel moment est-il souhaitable d'engager le dialogue avec l'utilisateur?
- Comment convient-il de formuler les questions?

Pour une traduction interactive

La première question que je voudrais aborder ici est celle de la justification du recours à une approche interactive au problème de la traduction par ordinateur. En effet, d'aucuns pourraient objecter qu'il n'y a guère d'intérêt à développer des systèmes qui exigent de la part de leurs utilisateurs un travail non négligeable et probablement très astreignant s'il existe par ailleurs des systèmes de traduction capables de fonctionner de façon autonome. Autrement dit, pourquoi recourir à un mécanisme semi-automatique alors qu'il en existe d'entièrement automatiques?

Il est généralement admis que pour satisfaire les attentes de ses utilisateurs, la traduction par ordinateur doit parvenir à un haut niveau de qualité et de fiabilité. Or, sur la base des connaissances actuelles et de la technologie disponible dans les domaines de la linguistique théorique, de la linguistique informatique et de la traduction, ces exigences ne sont pas réalistes, et il n'y a guère d'espoir de les voir réalisées dans un futur proche par un système autonome[1].

Le processus de traduction repose sur un large éventail de connaissances, parmi lesquelles figurent en première place les connaissances grammaticales de la langue source et de la langue cible. Il est clair, cependant, que ces connaissances linguistiques sont insuffisantes pour assurer une traduction correcte et de bon niveau. Il est en effet souvent indispensable de disposer soit de connaissances plus générales, de nature encyclopédique, soit encore de connaissances spécifiques au domaine concerné par le texte. Cependant, alors que des progrès considérables ont été accomplis dans certains champs de la linguistique théorique, et en particulier dans les domaines de la syntaxe et de la morphologie, domaines pour lesquels il est maintenant possible d'envisager des applications informatiques relativement sophistiquées, d'autres aspects du langage et du processus de traduction restent fort mal compris. C'est en particulier le cas de la sémantique et de la pragmatique, pour lesquelles nous ne disposons que de modèles

1. Ces remarques ne s'appliquent pas aux applications qui touchent des domaines très limités et très spécifiques, comme le système TAUM-Météo (*cf.* Isabelle 1987) qui traduit de façon autonome et satisfaisante des bulletins météorologiques, système que Slocum (1985 : 8) décrit à juste titre comme « the world's only example of a truly fully-automatic MT system ».

très partiels. Quant aux modèles computationnels de représentation de la connaissance, ils n'existent pour l'instant que pour des domaines microscopiques[2].

L'incapacité des systèmes de traduction automatique actuels à assurer une qualité et une fiabilité suffisantes est largement reconnue, et dans la pratique, ces systèmes de traduction exigent un travail non négligeable de correction (et/ou de préparation) de la part de leurs utilisateurs. Dans des systèmes conçus comme des systèmes autonomes, cette intervention humaine, reconnue indispensable, s'exerce par nécessité sous la forme de la préédition et de la postédition. Dans le premier cas, l'utilisateur vient en aide au système de traduction en tentant d'anticiper les problèmes les plus visibles. Dans le deuxième cas, l'utilisateur essaie de corriger les problèmes les plus criants dans le document traduit, un travail souvent ingrat et peu apprécié des traducteurs professionnels. L'intérêt de l'approche interactive est de permettre à l'intervention humaine dans le processus de traduction de s'exercer non pas avant ou après la traduction d'un texte, mais pendant celle-ci, c'est-à-dire au moment où un problème surgit.

Quelques avantages des systèmes interactifs

Parmi les avantages que présentent les systèmes de traduction interactifs par rapport à des systèmes autonomes, le plus évident est sans doute le fait que l'utilisateur conserve un certain contrôle sur le déroulement du processus de traduction. Ainsi, face à des problèmes délicats d'ambiguïté, le système peut afficher les diverses possibilités et laisser le soin à l'utilisateur de sélectionner la plus appropriée d'entre elles. En effet, il arrive souvent que la sélection de la possibilité la plus satisfaisante repose sur des critères faisant intervenir des connaissances non linguistiques (connaissances encyclopédiques, bon sens, etc.). Une telle démarche présente le double avantage 1) de libérer le système informatique d'une tâche difficile, qu'il ne pourrait accomplir de façon satisfaisante, et 2) d'avoir recours au système expert le plus fiable et le plus performant en matière de compréhension du langage, l'homme.

Il arrive couramment que les textes à traduire contiennent des expressions peu claires ou ambiguës, même pour un traducteur humain connaissant bien le domaine. Il se peut également que l'information linguistique du texte source soit sous-spécifiée, en ce sens que certains traits linguistiques pertinents pour la langue cible ne figurent pas dans le texte source. Cela est particulièrement fréquent au niveau lexical, comme l'illustre le cas de l'anglais *river* qui correspond parfois à *rivière*, parfois à *fleuve*. Il est possible — mais non certain — que le texte contienne certains indices permettant de choisir le terme le plus adéquat, mais la recherche de ces indices et leur interprétation est une tâche pour laquelle l'homme est sans conteste mieux équipé qu'un programme. Un cas particulièrement extrême de sous-spécification est souvent cité dans la littérature. Il s'agit du problème du choix des déterminants dans la traduction du japonais vers l'anglais. En effet, comme les syntagmes nominaux en japonais ne portent pas de

2. Voir Carbonnel et Tomita (1987) pour une présentation d'un prototype de système de traduction basé sur la connaissance.

marque défini/indéfini, le traducteur doit déterminer pour chaque syntagme nominal s'il convient de l'exprimer en anglais sous la forme d'un syntagme défini ou indéfini.

Pour résoudre des problèmes de ce type, un système autonome devrait avoir accès à un énorme réservoir de connaissances, être capable de raisonnements très élaborés, et finalement compter sur la postédition. En l'absence de systèmes généraux de connaissances et d'inférences, les systèmes de traduction autonomes n'ont d'autre solution que d'utiliser des paquets de règles et de traits *ad hoc*, ce qui a pour conséquence d'augmenter la complexité des systèmes et d'en réduire la fiabilité. Un système interactif n'aura pas besoin de se soucier de ce type de problèmes puisqu'il peut se reposer sur l'intervention de l'utilisateur. En règle générale, chaque fois qu'un problème surgit qui dépasse les capacités du système informatique, celui-ci engage un dialogue avec l'utilisateur, à qui il incombe de résoudre le problème. Il s'ensuit qu'un système interactif tend à être plus réduit et plus simple.

Un autre avantage que l'on peut attendre d'un système de traduction interactif est la fiabilité, qui découle du fait que l'utilisateur conserve tout au long du processus de traduction une certaine maîtrise sur le processus de décision. En particulier, le fait que les problèmes les plus délicats sont directement soumis à l'appréciation de l'utilisateur limite considérablement le risque de propagation d'erreurs ainsi que celui de l'effet combinatoire des ambiguïtés. On appelle propagation d'erreur le fait qu'une erreur commise à un certain niveau dans le processus de traduction est fréquemment susceptible d'entraîner des erreurs à d'autres niveaux. La chaîne d'erreurs ainsi créée peut avoir des conséquences extrêmement dommageables et implique souvent une réanalyse complète de la phrase (ou du paragraphe) dans la phase de postédition.

Sur la base des avantages décrits dans cette section, on peut se demander pourquoi une telle approche n'a pas été développée auparavant. En effet, il peut apparaître bien surprenant que cette idée souvent discutée dans des « papiers académiques » et notamment dans Kay (1980), Tomita (1984, 1986), Johnson & Whitelock (1987) et Melby (1982, 1987), et plus récemment dans Zajac (1988), Boitet (1990), Jones & Tsujii (1990) parmi beaucoup d'autres, n'ait guère suscité de développements à grande échelle, à l'exception peut-être du système ITS décrit par Melby (1980, 1982). On notera, par contre, que de nombreux systèmes d'aide à la traduction ont été développés sur la base d'approches interactives, mais ces systèmes se limitent habituellement à des tâches simples, telles que la consultation de dictionnaires ou d'encyclopédies, ou encore la correction orthographique.

La raison, sans doute, est que le développement d'un système de traduction interactif, bien que plus petit et plus simple au niveau de son système de règles qu'un système autonome, reste une entreprise extrêmement complexe. De plus, comme nous allons le voir, les systèmes interactifs imposent des contraintes sévères à l'architecture du système de traduction ainsi qu'à la nature des algorithmes.

Structure et organisation d'un système interactif

Dans la section précédente, nous avons discuté un certain nombre d'avantages que les systèmes de traductions interactifs présentent sur leurs rivaux non interactifs. Au

nombre de ces avantages figurent 1) une plus grande fiabilité, 2) une relative simplicité (de leurs règles), puisqu'ils n'ont pas besoin de *tout savoir*, et 3) un garde-fou à la propagation et à la prolifération des erreurs. Nous allons voir, maintenant, que ces avantages ne sont pas gratuits, et que les systèmes interactifs imposent certaines contraintes et restrictions.

Tout d'abord, par définition, un système interactif nécessite un utilisateur pour l'assister, c'est-à-dire répondre à ses requêtes. Il s'agit là d'une contrainte majeure, probablement incompatible avec certaines applications de traduction. Cependant, cette contrainte peut sembler moins « drastique » lorsqu'on la compare aux tâches de prédédition ou de postédition des systèmes autonomes.

Du point de vue de l'utilisateur, une exigence fondamentale est que le niveau d'interaction soit raisonnable. En d'autres termes, les dialogues devraient être limités à un minimum. Les utilisateurs, en effet, abandonneraient sans doute rapidement un système qui poserait des questions à chaque mot!

La question de savoir quel taux d'interaction les utilisateurs considéreront comme acceptable est délicate et ne peut être résolue hors contexte. Il est clair que des paramètres tels que la complexité du texte à traduire, la qualité des résultats, le contexte même de la traduction entrent en compte, de même que des facteurs personnels.

Ce qui est sûr, cependant, c'est que les questions non pertinentes doivent être évitées à tout prix. Or, cela n'est malheureusement pas forcément facile. Pour comprendre pourquoi, commençons par considérer la question fondamentale du choix du moment le plus opportun pour engager une interaction. Les exemples donnés en (1) permettront d'illustrer le problème :

(1)a. John knew the answer.
b. John knew the answer was wrong.

Dans la phrase (1a) le syntagme nominal *the answer* fonctionne comme objet direct du verbe *knew*, tandis que dans l'exemple (1b) le même syntagme correspond au sujet de la phrase enchâssée *the answer was wrong*. Avant d'atteindre le verbe enchâssé, l'analyseur syntaxique, qui parcourt la phrase de gauche à droite et qui sait que le verbe *to know* peut prendre soit un objet direct, soit une phrase enchâssée, pourrait être tenté de lever cette ambiguïté immédiatement et, par conséquent, d'interroger l'utilisateur pour lui demander quelle option est la bonne dans ce contexte. Une telle question, bien sûr, est inutile, puisque l'ambiguïté disparaît d'elle-même lorsque l'analyseur arrive à la fin de la phrase, dans l'exemple (1a), ou lorsque l'analyseur lit le verbe *was*, dans l'exemple (1b).

On observera qu'en général il y a un conflit entre les exigences de l'utilisateur et celles du système par rapport à cette question du moment opportun de la consultation de l'utilisateur. Du point de vue de ce dernier, moins il y aura de questions, mieux cela vaudra, ce qui signifie que l'interaction devrait être différée le plus possible, de façon à pouvoir éliminer toutes les questions inutiles. Du point de vue du bon fonctionnement du système, par contre, il est souhaitable que l'interaction prenne place le plus rapidement possible, puisque c'est l'intervention de l'utilisateur qui va permettre de sélectionner les solutions les plus pertinentes et d'éliminer les moins pertinentes.

Une intervention rapide dans un cas d'ambiguïté permet d'éviter la propagation des erreurs et la combinaison des solutions.

À un autre niveau, il convient d'insister sur la nécessité d'un dialogue aussi convivial que possible. En particulier, il est important que les questions soient formulées de telle façon qu'un utilisateur non spécialiste de la théorie linguistique sous-jacente au système soit néanmoins capable de les comprendre. L'utilisation de fenêtres de dialogue et la présentation des diverses possibilités sous forme de menus dans lesquels l'utilisateur choisit une ou plusieurs réponses semblent répondre à cette attente.

Finalement, comme pour tous les systèmes qui travaillent en temps réel, il y a une contrainte d'efficacité. Il est en effet fondamental que les dialogues se suivent rapidement, de telle façon que l'utilisateur ne s'impatiente pas ou qu'il ne relâche pas son attention.

Il est intéressant de noter que ces contraintes ont des conséquences importantes sur l'architecture du système de traduction ainsi que sur le choix des algorithmes. Pour illustrer ce point, que nous ne pourrons développer en détail ici, bornons-nous à la question du moment de l'interaction dans la phase d'analyse d'une phrase. Il est clair que pour poser des questions aussi pertinentes que possible, l'analyseur doit non seulement noter la présence d'alternatives à un moment donné, mais également explorer dans une certaine mesure ces différentes possibilités, de façon à pouvoir garantir qu'elles ont une certaine pertinence dans le contexte en question. Comme, par ailleurs, il n'est pas souhaitable de différer toutes les questions jusqu'à ce que l'analyseur ait atteint la fin de la phrase, un certain degré de parallélisme dans le traitement des solutions semble nécessaire. De même, il apparaît que les deux stratégies classiques, ascendante et descendante, ne sont guère satisfaisantes et qu'une stratégie mixte semble préférable si l'on veut éviter les longues séquences d'hypothèses qui se révèlent finalement incompatibles avec les mots de la phrase à analyser (stratégie descendante) ou les nombreuses structures localement bien formées mais incompatibles avec leur contexte gauche (stratégie ascendante).

Niveaux d'interaction

Afin de rendre cette discussion un peu plus concrète, nous allons considérer dans cette section le cas d'un hypothétique système de traduction interactif et montrer les différents niveaux d'interaction. Nous admettrons que le système repose sur l'architecture classique avec les trois composantes d'analyse, de transfert et de génération. De plus, nous conviendrons que toute l'interaction est limitée aux deux premières composantes et, par conséquent, à une clarification du texte source.

Un premier niveau d'interaction prend place lorsque le texte source est considéré par l'analyseur, lors de l'analyse lexicographique. Le programme peut rencontrer des mots qui lui sont inconnus, c'est-à-dire des mots qui ne figurent pas dans ses dictionnaires. Il peut s'agir de noms propres ou de noms communs qui ne se trouvent pas dans ses dictionnaires, ou encore de mots mal orthographiés. Une fenêtre d'interaction est alors ouverte, exposant la phrase et mettant en évidence le ou les mots concernés.

Un menu permet à l'utilisateur de choisir l'action la plus appropriée parmi les possibilités suivantes :

- corriger la phrase;
- valider le mot comme nom propre;
- créer une nouvelle entrée de dictionnaire;
- interrompre le processus.

Un deuxième niveau d'interaction survient lors de l'analyse syntaxique, lorsque l'analyseur rencontre des ambiguïtés qu'il ne parvient pas à surmonter au moyen de ses règles et heuristiques. Cela concerne les cas typiques d'ambiguïtés syntaxiques, telles que les problèmes d'attachement illustrés en (2) :

(2)a. Jean frappe le chien avec le collier.
b. Jean frappe le chien avec les taches blanches.
c. Jean frappe le chien avec le poing.

Dans la phrase (2a), le syntagme prépositionnel *avec le collier* peut se rattacher soit au syntagme nominal *le chien* et alors il est interprété comme un modificateur de ce syntagme nominal, soit au verbe *frapper* et alors il est interprété comme un complément de ce verbe. La première interprétation est sans doute la plus naturelle dans le cas de la phrase (2b), la deuxième dans la phrase (2c). Dans des cas de ce type, la fenêtre de dialogue peut afficher la chaîne de caractères correspondant au syntagme prépositionnel à attacher et les têtes des syntagmes auquel il peut s'attacher. Dans notre exemple, cela reviendrait à afficher le syntagme prépositionnel [$_{PP}$ avec le collier] et demander s'il faut l'attacher au syntagme nominal objet direct *chien* ou au verbe *frappe*. Comme toujours, l'utilisateur a également la possibilité de modifier la phrase ou d'interrompre la session.

Un deuxième exemple d'ambiguïté syntaxique courante concerne l'interprétation de syntagmes antéposés, comme les syntagmes interrogatifs. Ainsi, dans la phrase (3) ci-dessous, *à qui* peut être interprété soit comme un complément du verbe *promettre* (*cf. Jean a promis à qui d'écrire*), soit comme un complément du verbe enchâssé *écrire*. La fenêtre d'interaction, dans un cas de ce genre, propose les possibilités, ainsi formulées :

(3) À qui avez-vous promis d'écrire?
a. écrire à qui?
b. promettre à qui?

Considérons un troisième cas classique d'ambiguïté syntaxique, cette fois en anglais.

(4)a. Flying planes can be dangerous.
b. Amusing people can be fun.
c. Visiting relatives can be boring.

Dans les phrases (4), les formes en *ing* (*flying, amusing, visiting*) peuvent être interprétées comme des adjectifs prénominaux modifiant respectivement les noms *planes, people, relatives*, ou comme des gérondifs, auquel cas les noms fonctionnent

commes des objets directs. La fenêtre de dialogue pourrait afficher les formes en *-ing* et les deux types d'interprétations, ce qui donnerait pour la phrase (4b) :

- *amusing* comme modificateur de [$_{NP}$ people];
- *amusing* comme verbe avec [$_{NP}$ people] comme complément.

Un troisième niveau d'interaction concerne les ambiguïtés d'ordre sémantique ou pragmatique, comme dans les exemples suivants :

(5)a. Jean ne boit pas parce qu'il est triste.
b. Jean dit à Paul que son livre se vend bien.
c. Les enfants se sont lavés.

La phrase (5a) illustre le problème des ambiguïtés de champ (ou de portée), c'est-à-dire du domaine sur lequel porte un quantificateur ou une négation. Ainsi, dans cette phrase, la négation peut porter soit sur le verbe seul, soit sur le verbe et la phrase circonstancielle. Dans le premier cas on obtient l'interprétation correspondant à la paraphrase *c'est parce qu'il est triste que Jean ne boit pas*, dans le second l'interprétation, très différente, correspondant à *ce n'est pas parce qu'il est triste que Jean boit*.

Dans de tels cas, il serait tentant d'afficher dans la fenêtre de dialogue les paraphrases correspondant aux diverses interprétations et de demander à l'utilisateur de choisir la plus appropriée. Cependant, le recours à des paraphrases peut être difficile et parfois conduire à de nouvelles ambiguïtés. Il est donc préférable dans de tels cas d'éviter les paraphrases et de formuler les questions comme suit :

- *parce qu'il est triste* est dans le domaine de la négation;
- *parce qu'il est triste* n'est pas dans le domaine de la négation.

Le deuxième exemple d'ambiguïté non syntaxique concerne l'interprétation des pronoms, comme dans l'exemple (5b). Dans cette phrase, le possessif *son* peut renvoyer soit à *Paul*, soit à *Jean*, soit encore à une autre personne mentionnée précédemment dans le discours. Ici, à nouveau, la fenêtre d'interaction va afficher la liste des antécédents potentiels et demander à l'utilisateur d'effectuer le bon choix.

Enfin, dans le cas de la phrase (5c), l'ambiguïté concerne l'interprétation du pronom anaphorique *se*. Dans cette phrase, *se* peut être compris avec un sens réfléchi ou réciproque. La fenêtre d'interaction peut afficher l'information suivante :

(6) Les enfants se sont lavés.
a. les uns les autres? (réciproque)
b. eux-mêmes? (réfléchi)

De nombreuses ambiguïtés se présentent au niveau du transfert, et notamment du transfert lexical, lorsque le système doit décider comment un élément lexical de la langue source doit être traduit. On sait que la correspondance entre items lexicaux de deux langues n'est pratiquement jamais une correspondance univoque. Autrement dit, il y a pratiquement toujours plusieurs possibilités de traduction pour un mot donné, même lorsque l'environnement syntaxique de ce mot a été bien dégagé. Ceci est en partie une conséquence directe de l'ambiguïté lexicale, dans la mesure où les différents sens associés à un mot sont susceptibles de traductions différentes. Il est légitime de penser que le dictionnaire bilingue contient toutes les correspondances possibles pour un item donné. On peut donc afficher cette liste dans la fenêtre d'interaction. Cette

façon de faire nécessite, cependant, une connaissance de la langue cible. Si, comme nous le souhaitons, l'interaction peut être limitée à la langue source, la sélection d'une correspondance se fera sur la base des différents sens du mot dans la langue source. On affichera alors le mot avec ses différents sens, en utilisant des synonymes. Par exemple, dans le cas (très simplifié) du mot *assurance*, on affichera:

- *assurance* syn. confiance;
- *assurance* syn. contrat.

Considérons un deuxième exemple, repris de Boitet (1990). Il s'agit de la traduction du mot *mur* dans une langue (comme l'espagnol) qui utilise des mots distincts pour le mur intérieur et le mur extérieur. La fenêtre de dialogue pourrait alors prendre l'allure suivante :

- *mur* — intérieur;
- *mur* — extérieur;
- ...

La sélection du sens le plus approprié d'un mot, qui permet au système de choisir une traduction correcte, est basée sur des exemples ou des synonymes associés aux entrées lexicales.

Conclusion

Les approches interactives au problème de la traduction par ordinateur présentent une solution de rechange particulièrement séduisante. En associant l'utilisateur au processus de traduction, les systèmes interactifs peuvent tirer parti des avantages respectifs de l'homme et de la machine. En outre, le fait que l'utilisateur conserve un certain contrôle sur le fonctionnement augmente la fiabilité et la robustesse du système. Enfin, alors que les systèmes autonomes nécessitent un long travail de postédition, tâche particulièrement délicate qui nécessite les compétences d'un traducteur professionnel, les systèmes interactifs du type discuté ici limitent l'interaction entre l'utilisateur et le programme à des dialogues en temps réel, habituellement sous la forme de menus dans lesquels l'utilisateur choisit un ou plusieurs items. De plus, si, comme nous l'admettons, des phrases complètement désambiguïsées peuvent être correctement traduites de façon automatique, il devrait être possible, en principe, de restreindre les interactions aux composantes d'analyse et de transfert et, par conséquent, de restreindre le dialogue à la seule langue source. La perspective de systèmes de traduction qui ne requièrent pas une connaissance de la langue cible n'est sans doute pas le moindre des avantages d'une telle approche.

Bibliographie

BOITET C. (1990) « Towards Personal MT », *Coling-90*, vol 3, p. 30-35.

CARBONELL J. and M. TOMITA (1987) « The Universal Parser Architecture for Knowledge-Based Machine Translation », *Proceedings of the 10th International Joint Conference on Artificial Intelligence*.

ISABELLE P. (1987) « Machine Translation at the TAUM Group », M. King (éd.), *Machine Translation Today*, Edinburgh University Press.

JONES D. and J. TSUJII (1990) « Interactive High-Quality Machine Translation for Monolinguals », *Proceedings of the 3rd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Linguistic Research Center, Austin University of Texas.

JOHNSON R. and P. WHITELOCK (1987) « Machine Translation as Expert Task », S. Nirenburg (éd.), *Machine Translation : Theoretical and Methodological Issues*, Cambridge University Press.

KAY M. (1980) *The Proper Place of Men and Machines in Language Translation*, Technical report CSL-80-11, Xerox PARC.

KAY M. (1982) « Machine Translation », *American Journal of Computational Linguistics*, 8 : 74-78.

MELBY A., M. SMITH and J. PETERSON (1980) « ITS : Interactive Translation System », *Proceedings of the 8th International Conference on Computational Linguistics (COLING-80)*.

MELBY A. (1982) « Multi-level translation aids in a distributed system », *Proceedings of the 9th International Conference on Computational Linguistics (COLING-82)*, North-Holland.

MELBY A. (1987) « On Human-Machine Interaction in Translation », S. Nirenburg (éd.), *Machine Translation : Theoretical and Methodological Issues*, Cambridge University Press.

SLOCUM J. (1985) « A Survey of Machine Translation : its History, Current Status, and Future Prospects », *Computational Linguistics*, 11 : 1-17.

SLOCUM J. (éd.) (1988) *Machine Translation Systems*, Cambridge University Press.

TOMITA M. (1984) « Disambiguating Grammatically Ambiguous Sentences by Asking », *Proceedings of the 10th International Conference on Computational Linguistics (COLING-84)*.

TOMITA M. (1986) *Efficient Parsing for Natural Language*, Kluwer Academic Publishers.

ZAJAC, R. (1988) « Interactive Translation : a New Approach », *Proceedings of the 12th International Conference on Computational Linguistics (COLING-88)*.

7

La traduction automatique des bulletins d'avalanches de la Suisse

Pierrette BOUILLON, Katharina BOESEFELDT

ISSCO, Université de Genève, Genève, Suisse

Introduction

L'objectif de cette présentation est de montrer comment le système ELU (*Environnement Linguistique d'Unification*), développé à l'ISSCO, est utilisé dans le cadre concret de la construction d'un système de traduction automatique portant sur un domaine limité. Dans l'exposé, nous mettrons particulièrement l'accent sur la notion de représentation formelle et nous montrerons comment une riche représentation sémantique en structures d'attributs permet un transfert plus efficace entre les langues, ainsi qu'une approche moderne de la traduction automatique.

Après une brève présentation du projet, nous aborderons la mise en œuvre informatique avec le logiciel ELU, puis nous donnerons quelques exemples de représentation formelle en structures d'attributs.

Description du projet de traduction automatique des bulletins d'avalanches

Le projet

Le projet a pour but l'implémentation d'un système de traduction automatique capable de traduire des bulletins d'avalanches de la Suisse émis par *l'Institut Fédéral Suisse*

pour l'Étude de la Neige et des Avalanches (IFENA)[1] une à plusieurs fois par semaine pendant l'hiver. Rédigés initialement en allemand, ces bulletins doivent être traduits dans les deux autres langues officielles : le français et l'italien. Leur traduction automatique devrait permettre une traduction plus rapide, plus cohérente et plus systématique.

Le monde des avalanches : un sous-langage

Les bulletins d'avalanches constituent une application idéale pour la traduction automatique. Ils traitent du domaine limité des avalanches en Suisse. Ce domaine est bien connu et bien formalisé : des spécialistes ont défini les lois générales météorologiques, la manière de les exprimer et les zones géographiques de la Suisse (Salm 1986). Il est donc possible de définir le sous-langage utilisé pour la rédaction des bulletins d'avalanches et de limiter le traitement automatique à ce sous-langage. Or, dans un sous-langage, les ambiguïtés lexicales, syntaxiques et sémantiques sont réduites et plus aisément traitées. D'une part, le nombre de mots ainsi que leur sens sont restreints et les mêmes expressions reviennent à plusieurs reprises. D'autre part, les structures syntaxiques sont en nombre limité puisqu'il n'est pas nécessaire et même peu souhaitable pour des raisons de compréhension d'exprimer la même idée de différentes manières (Lehrberger 1982).

Unité sémantique des bulletins d'avalanches

Le but des bulletins d'avalanches est de donner une brève description d'ensemble des conditions de neige et de décrire le degré régional de danger actuel et prévu dans les Alpes suisses. Les bulletins sont généralement articulés en trois alinéas : la situation générale, les degrés du danger et l'indication des mesures à prendre et des dangers particuliers (annexe 1).

Chacune de ces trois parties utilise un vocabulaire bien défini et les bulletins sont donc rédigés au moyen d'un nombre limité de mots (environ mille), qui revêtent un nombre fini de significations et sont utilisés dans un nombre restreint de constructions. Pour vérifier la cohérence des syntagmes nominaux, il est assez aisé de définir un ensemble de traits sémantiques, de coder les noms et adjectifs avec ces traits et de stipuler que les mêmes traits doivent être récurrents à travers le syntagme. Le groupe nominal *le grand danger général*, par exemple, est correct, car le nom *danger* et les adjectifs *grand* et *général* ont reçu le même type sémantique **danger**.

De même, il est possible d'établir les unités de sens qui constituent les concepts de base du monde des avalanches et de les traiter comme des entités globales. Pour traiter, par exemple, le syntagme nominal français *couverture de neige* qui constitue un seul concept et qui, sémantiquement, doit être considéré comme une unité de sens,

1. Nous remercions l'IFENA qui subventionne partiellement le projet.

nous stipulons dans le lexique que le nom *couverture* requiert un argument qui a pour valeur *neige*.

Unité syntaxique des bulletins d'avalanches

La syntaxe, bien qu'elle soit, elle aussi, limitée, présente toutefois un degré de complexité assez intéressant, qui nous oblige à aborder des problèmes linguistiques parfois peu étudiés.

Contrairement aux notices météorologiques de TAUM-METEO (Lehrberger 1982), les bulletins sont en effet écrits à la troisième personne en prose et incluent des phrases simples ou subordonnées actives, passives, positives et négatives au présent, passé et futur, parfois coordonnées. Les groupes nominaux, adjectivaux, prépositionnels et verbaux sont souvent très complexes, avec plusieurs niveaux d'enchâssement, et ils sont souvent coordonnés.

Mise en œuvre du projet avec ELU : un environnement linguistique d'unification pour construire des systèmes de traduction automatique

Après avoir présenté le logiciel ELU, nous verrons la méthode de traduction qu'il utilise, puis nous mettrons en évidence l'originalité de ce logiciel. Enfin, nous aborderons l'état d'avancement du projet.

Le logiciel ELU

ELU (Estival 1989 et 1990, Russell 1990) est un programme qui a été développé à l'ISSCO et qui est capable d'interpréter des grammaires écrites dans un langage particulier. Inspiré de PATR II (Shieber 1986), il utilise le principe de l'unification pour combiner les informations, des ensembles de traits et de valeurs appelés structures d'attributs (voir exemple 1). Il possède trois modules principaux : **l'analyseur, le générateur et le module de transfert**.

L'analyseur permet d'obtenir pour une phrase les représentations en structures d'attributs permises par la grammaire.

Le générateur part d'une représentation de ce type et applique les règles de la grammaire pour découvrir la ou les chaînes de caractères qui peuvent être associées à la représentation en question.

Le module de transfert établit une relation entre deux mots ou deux représentations en structures de traits, pour permettre le passage d'une langue à une autre. Par exemple, la règle de transfert lexicale suivante met en correspondance directe les mots *Gefahr* et *danger* :

: TA : Gefahr danger

La règle de transfert structurale **pred** établit, elle, un lien entre les deux structures de traits < *head sem pred* > = *Valeur* de deux langues différentes L1 et L2 et lie la valeur X1 du chemin <* *head sem pred* > d'une langue L1 et la valeur X2 du même chemin dans l'autre langue L2 :

:T : pred
: L1 : <* head sem pred > = X1
: L2 : <* head sem pred > = X2
: X : X1 = X2

La traduction automatique avec ELU

La traduction avec ELU, fondée sur le modèle du transfert, se fait donc en trois étapes : la phrase en langue source est d'abord analysée pour produire une représentation en structures d'attributs; cette structure d'attributs est ensuite transformée par les règles de transfert dans une représentation conforme à la langue cible. Enfin, le générateur part de cette représentation pour générer le texte en langue cible.

Originalité de l'approche avec ELU

L'originalité de l'approche avec ELU réside dans la richesse de sa représentation en structures d'attributs, une génération et un transfert basés uniquement sur la représentation sémantique, la possibilité d'utiliser la même grammaire pour l'analyse et la génération, ainsi qu'un transfert bidirectionnel.

Représentation en structures d'attributs

Les ensembles de traits et de valeurs, appelés structures d'attributs (voir exemple 1), offrent un moyen flexible, élégant et efficace pour exprimer, de manière unifiée et condensée, toutes les informations jugées pertinentes par le linguiste, que ces dernières soient morphologiques, sémantiques ou syntaxiques.

Une génération et un transfert basés sur la représentation sémantique en structures d'attributs

Dans notre projet, la génération et le transfert se font uniquement sur la base d'une représentation sémantique, la syntaxe et les informations morphologiques non pertinentes pour la traduction et spécifiques aux langues n'étant pas prises en considération. Dans la sémantique, nous stipulons seulement quel est le prédicat de phrase, son ou ses arguments (dans une liste), le temps (présent, passé ou futur), si la phrase est positive ou négative et quels sont les modificateurs (voir exemple 1).

Utilisation de la même grammaire pour l'analyse et la génération

ELU permet d'utiliser la même grammaire pour l'analyse et la génération. Si l'utilisation d'une seule grammaire rend parfois plus complexe la tâche de celui qui écrit la grammaire, ce système offre cependant des avantages non négligeables. Tout d'abord, il ne nécessite pas l'écriture, puis le test fastidieux de deux grammaires différentes. De plus, il offre un moyen idéal pour vérifier si la grammaire écrite n'accepte pas des phrases illicites. Enfin, il assure la cohérence du système : toutes les phrases analysées peuvent aussi être générées par le système.

Un transfert bidirectionnel

Un des postulats d'ELU, qui découle de sa conception, est la possibilité de rendre le transfert réversible. S'il est possible de passer de la structure A à la structure B, étant donné la règle de transfert X, le contraire devrait aussi pouvoir être vérifié. En suivant cet objectif, nous pouvons donc tester les conséquences de cette approche moderne qui nous offre la possibilité d'être cohérents.

État d'avancement du projet

Après une étude fine du corpus avec l'aide de l'IFENA, l'accent est mis sur la traduction. Deux lexiques et deux grammaires du français et de l'allemand sont en cours d'élaboration et traitent de tous les phénomèmes de base du corpus des avalanches. Elles permettent l'analyse et la génération d'un grand nombre de phrases du corpus de complexité variable (annexe II). En même temps, nous rédigeons les règles de transfert qui permettent déjà la traduction bidirectionnelle de phrases du type de celles citées en annexe III. C'est de la représentation formelle des phrases qui figurent dans les annexes I et II qu'il sera question dans le chapitre cinq.

Exemples de représentation formelle des phrases en structures d'attributs

Comme nous l'avons dit plus haut, les phrases sont représentées par des structures d'attributs. Comme le monde des avalanches constitue un sous-langage où les variations structurales entre les langues sont peu nombreuses, nous avons essayé, lorsque c'était possible, de représenter les phrases allemandes et leur traduction française avec les mêmes structures sémantiques. Nous complexifions ainsi les grammaires d'analyse, mais nous simplifions considérablement les règles de transfert et nous gagnons ainsi en efficacité. Le transfert, en effet, est plus expérimental, encore moins bien connu et gagne à être limité.

Les exemples d'analyse suivants montrent la force d'expression des structures d'attributs, qui permettent d'exprimer le contenu sémantique des phrases. D'une part,

les structures de listes ([A,B]) et d'arbres (X(Y)) permettent une représentation très compacte. D'autre part, l'utilisation de structures rentrantes où deux traits possèdent la même valeur (#41 dans l'exemple 2) permet d'exprimer aisément des faits linguistiques complexes.

Exemple 1 (phrase active) :

in der Westschweiz besteht oberhalb 2000 m eine mässige lokale Schneebrettgefahr, traduit automatiquement par *un danger local modéré de plaques de neige subsiste au-dessus de 2000 m en Suisse Romande.*

```
args = [⟨81⟩ head : sem : compl : head : sem : compl : head : sem : compl = []
                                                                detype = definite
                                                                mod : n-comp : head : sem : pred = []
                                                                      nom = []
                                                                pred = Schnee
                                                detype = non
                                                mod : n-comp : head : sem : pred = []
                                                      nom = []
                                                pred = brett
                                    detype = indefinite
                                    mod : alt : head : sem : pred = []
                                          expo : head : sem : pred = []
                                          n-comp : head : sem : pred = []
                                          nom : [⟨601⟩ head : sem : mod : degre : head : sem : pred = []
                                                                          pred = mässig
                                                ⟨602⟩ head : sem : mod : degre : head : sem : pred = []
                                                                          pred = lokal
                                          rel = []
                                    pred = gefahr ]
mod : alt : head : sem : mod : degre : head : sem : pred = []
                         pred = oberhalb (m(2000))
      loc : head : sem : compl = []
                         detype = definite
                         mod : expo : head : sem : pred = []
                               n-comp : head : sem : pred = West
                               nom = []
                               rel = []
                         pred = schweiz
      temps = []
morph : temps = present
        voix = actif
positif = yes
pred = bestehen
```

Cette structure d'attributs spécifie que la phrase est positive et au présent actif. La phrase possède comme prédicat le verbe *bestehen*, comme argument le groupe nominal sujet *eine mässige lokale Schneebrettgefahr* et comme modificateurs les groupes prépositionnels d'altitude et de lieu. Pour indiquer qu'il n'y a pas de groupe

prépositionnel de temps, nous utilisons une liste vide []. Pour que le transfert entre le français et l'allemand soit généralisable, nous avons décidé de décomposer les mots composés allemands dans le lexique pour obtenir la même représentation en allemand et en français. *Schneebrettgefahr*, par exemple, est décomposé en trois composants : une tête *Gefahr* avec un complément complexe composé d'une tête *Brett* et d'un complément *Schnee*.

Exemple 2 (phrase prédicative) :

in den Alpen ist die Gefahr noch gering und lokal, traduit automatiquement en français par *le danger est encore local et modéré dans les Alpes*.

```
args = [⟨238⟩ head : sem : args = [⟨38⟩ head : sem : ⟨41⟩
                                                      compl = []
                                                      detype = definite
                                                      mod : alt : head : sem : pred = []
                                                            expo : head : sem : pred = []
                                                            n-comp : head : sem : pred = []
                                                            nom = []
                                                            rel = []
                                                      pred = Gefahr]
                                 mod : degre : head : sem : pred = []
                                 pred = gering
                 ⟨245⟩ head : sem : args = [⟨51⟩ head : sem  => #41]
                                 mod : degre : head : sem : pred = []
                                 pred = lokal]
mod : alt : head : sem : pred = []
      loc : head : sem : compl = []
                         detype = definite
                         mod : alt : head : sem : pred = []
                               expo : head : sem : pred = []
                               n-comp : head : sem : pred = []
                               nom = []
                               rel = []
                         pred = Alpen
      phrase : head : sem : pred = noch
      temps = []
morph : temps = present
        voix = actif
positif = yes
pred = und
```

Cette analyse montre un exemple d'utilisation de structures rentrantes pour exprimer le fait linguistique que les deux adjectifs coordonnés *gering und lokal* se rapportent au même nom *Gefahr* (#41). Dans les phrases prédicatives, le verbe *être* qui n'a pas de valeur sémantique n'apparaît pas dans la représentation. Le prédicat de la phrase qui est l'opérateur *und* possède deux arguments *gering* et *lokal* qui renvoient au même argument *Gefahr*.

Conclusion

Cet exposé montre une application concrète du logiciel d'ELU (*Environnement Linguistique d'Unification*). Après une brève description du projet, nous avons mis en évidence l'importance de la notion de sous-langage pour le développement d'un système de traduction automatique. La limitation du domaine nous a en effet permis la définition d'une représentation sémantique cohérente de la phrase pour un transfert efficace entre le français et l'allemand. Nous avons aussi montré que notre système de traduction automatique bénéficie de tous les avantages d'une approche nouvelle basée sur l'unification.

Bibliographie

ESTIVAL D. (1990a) *Elu User Manual, Technical Note 1*, ISSCO, Geneva.

ESTIVAL E., BALLIM A., RUSSELL G. and WARWICK S. (1990b) « A Syntax and Semantics for Feature-Structure Transfer », *The Third International Conference on Theorical and Methodological Issues in Machine Translation of Natural Language*, 11-13 juin 1990.

LEHRBERGER J. (1982) « Automatic Translation and the Concept of Sublanguage », *Sublanguage. Studies of Language in Restricted Semantic Domains*, edited by Richard Kittredge and John Lehrberger, Berlin, New York, de Gruyter, p. 81-106.

RUSSELL G., BALLIM A., ESTIVAL D. and WARWICK S. (1991) « A Language for the Statement of Binary Relation over Feature Structures », to appear in *Proceedings of European Association for Computational Linguistics*.

RUSSELL G., CARROLL J. and WARWICK S. (1990) « Asymmetry in Parsing and Generating with Unification Grammars : Case Studies from ELU », *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, June 1990, p. 205-211.

SALM B. (1982) *Lawinenkunde für den Praktiker*, Bern, Verlag des SAC.

SHIEBER S.M. (1986) « An Introduction to Unification Based Grammar », *CSLI Lecture Note*, n° 4.

Annexe I : Exemple de bulletin (9 mars 1984)

Starke Sonneneinstrahlung und eine Erwärmung zu Beginn der Woche sowie die nachfolgende kräftige Abkühlung haben die Lawinengefahr weiter gemindert. Die Niederschläge der letzten zwei Tage brachten lediglich dem Alpennordhang 10 bis 30 cm Schnee.

Am Alpennordhang, im Wallis in den Tessiner Bergen, in Nord- und Mittelbünden sowie im Unterengadin ist die Lawinengefahr dank einer stabilen Altschneedecke gering. Einzelne Gefahrenstellen können noch an Schattenhängen oberhalb rund 2000 m angetroffen werden.

Im Oberengadin und in den südlichen Bündner Tälern, wo das Schneedeckenfundament immer noch eine ungenügende Festigkeit aufweist, besteht für den Skifahrer vor allem an Nord- und Osthängen oberhalb rund 2000 m eine mässige örtlich beschränkte Lawinengefahr.

Annexe II : Exemples de phrases analysées et générées par la grammaire française

- Vingt à trente cm de neige fraîche sont tombés dans les Alpes, dans la région du Simplon, dans les vallées de la Viège et dans la région de la Maloja.
- Le rayonnement intense et l'élévation modérée de la température ont influencé le tassement de la couverture de neige.
- Les journées ensoleillées, mais froides ont influencé la stabilité de la couverture de neige.
- Les chutes de neige qui continuent causent une nouvelle aggravation du danger de avalanches.
- Une constitution défavorable de la couverture de neige incite à une prudence accrue sur les pentes raides.
- La couverture de neige présente des couches défavorables de givre qui incitent à la prudence.
- Le danger de plaques de neige est local et modéré dans les autres régions.
- Le touriste doit prendre garde à un danger élevé d'avalanches sur les pentes raides et ensoleillées où la couverture de neige est ramollie.
- Il faut considérer que la neige fraîche empêche le refroidissement de l'ancienne couverture humide.
- Bien que les chutes de neige soient insignifiantes, le danger local de plaques de neige augmente.
- Après que des avalanches sont descendues et qu'un refroidissement marqué s'est produit, la situation s'est calmée.
- Un danger local modéré de plaques de neige subsiste au-dessus d'environ 2000 m sur le versant nord des Alpes, dans la région du Gotthard, dans les Grisons et en Basse-Engadine.
- Le skieur doit éviter les pentes raides d'exposition nord à est.
- Les voies de communication sont menacées par des glissements de neige dans les bassins d'exposition sud.
- Les endroits dangereux se trouvent sur les pentes raides et sur les cuvettes d'exposition nord-ouest à nord-est.
- L'accès aux voies de communication et aux bâtiments est libéré.
- Après le temps variable des derniers jours, les hauteurs de neige atteignent environ 30 à 50 cm au-dessus de 1000 m sur le versant nord de les Alpes, en Engadine et en Valais après une semaine pauvre en précipitations.

Annexe III : Exemples de phrases transférées de l'allemand au français et du français à l'allemand

- *In den Alpen, in dem Simplongebiet, in den Vispertälern und in dem Malojagebiet fielen 20 bis 30 cm Neuschnee.*
- *Die Schneedecke betrug weniger als 10 cm.*
- *In den übrigen Gebieten beträgt der Neuschnee weniger als 10 cm.*
- *An dem Alpensüdhang fielen rund 80 cm Schnee.*

- *An dem ganzen Alpennordhang, in dem Wallis, in dem Gotthardgebiet und in dem Unterengadin besteht eine sehr grosse allgemeine Lawinengefahr.*
- *In dem Engadin besteht eine sehr grosse allgemeine Lawinengefahr.*
- *In den Alpen ist die Lawinengefahr noch gering und lokal.*
- *In den übrigen Gebieten ist die Schneebrettgefahr mässig und lokal.*
- *Der kräftige Temperaturanstieg und die nachfolgende Abkühlung beeinflussen die Schneedecke.*
- *An dem Alpennordhang und in dem Wallis betragen die Schneehöhen rund 20 bis 50 cm.*
- *Ein ungünstiger Schneedeckenaufbau ermahnt zu erhöhter Vorsicht.*
- *An dem Alpennordhang, in dem Wallis, in dem Gotthardgebiet und in dem Engadin besteht oberhalb 2000 m eine mässige lokale Schneebrettgefahr.*
- *In den übrigen Gebieten herrscht oberhalb 1800 m eine mässige lokale Schneebrettgefahr.*
- *Der wechselhafte Temperaturverlauf fu5hrte zu einer weiteren günstigen Setzung und Verfestigung der Schneedecke.*
- *Ein starker Temperaturanstieg und windbegleitete Schneefälle haben im Wallis zu einer Verscha5rfung der Lawinengefahr gefu5hrt.*
- *Die starke Erwärmung und die leichtere Abkühlung führten zu einer günstigen Setzung und Verfestigung der Schneedecke.*

8

Le projet EUROTRA, projet européen de recherche sur la traduction automatique

Christiane JADELOT, Dominique BOUSSEL

CELTA, CNRS / EUROTRA F-Nancy, Université Nancy 2, Nancy, France

EUROTRA

Finalité/originalité

Pour faire face à l'immense difficulté de traduction que représente le multilinguisme de la CEE (en même temps sa richesse), la Commission a lancé un projet de traduction automatique.

La Communauté européenne compte en effet 9 langues officielles : le danois, l'espagnol, l'allemand, le grec, l'anglais, l'italien, le néerlandais, le portugais et le français, ce qui fait 72 paires de langues si l'on veut les traduire toutes l'une vers l'autre.

Actuellement un certain nombre de documents ne sont pas traduits étant donné le coût très élevé du multilinguisme.

Face à cette situation la Commission a donc décidé de développer des outils d'aide à la traduction, le projet EUROTRA en fait partie.

Un prototype

Il s'agit en fait d'une maquette capable de traiter toutes les langues communautaires, et considérée comme une action de stimulation de la recherche visant à créer une

masse de connaissances en matière de traduction automatique (TA) et de linguistique informatique.

Au terme du programme on pourrait disposer d'un système opérationnel dans un domaine limité pouvant être développé ultérieurement dans la phase industrielle du projet.

Le système devrait couvrir des dictionnaires de 20 000 entrées pour chaque langue en 1992.

C'est en fait le premier programme communautaire de recherche (200 chercheurs) et de développement dans le domaine du traitement des langues naturelles contribuant :

- à la diffusion des connaissances et du savoir-faire,
- à la création d'infrastructures de recherche dans les pays membres,
- à la coopération européenne et à la prise de conscience d'une recherche communautaire dans ce domaine aussi importante que les objectifs techniques.

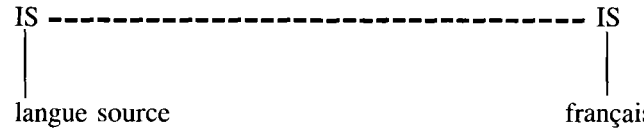
Le processus de traduction

Il se décompose en trois phases : l'analyse, le transfert et la synthèse.

L'analyse

Dans EUROTRA l'analyse, ainsi que la génération (synthèse) d'ailleurs, est monolingue. Elle est la même quelle que soit la langue cible. Cette caractéristique est motivée par le fait que le système traite un grand nombre de langues.

Il y a 9 modules d'analyse et 72 modules de transfert.



L'analyse d'une phrase repose sur une approche stratificationnelle très classique. Elle s'effectue selon quatre étapes :

1. analyse morphologique **EMS** : chaque mot de la phrase est étiqueté selon sa catégorie de discours;
2. analyse syntaxique **ECS** : les catégories du discours sont organisées en constituants tels que groupe nominal GN, groupe verbal GV, groupe prépositionnel GP;
3. analyse relationnelle **ERS** : les constituants sont réorganisés selon leur fonction syntaxique : verbe, sujet, objet, modifieur;

4. analyse sémantique **IS** : les fonctions syntaxiques sont organisées selon leur rôle sémantique : prédicat (gov), argument (arg) et modificateur (mod).

Chaque niveau d'analyse produit un arbre décoré. Le passage d'un niveau à l'autre est assuré par un transducteur. Le transfert d'un niveau à l'autre consiste en une modification d'arbres (ou élimination). Du point de vue du « framework » il n'y a pas de différence entre un transducteur du module monolingue et le transfert d'IS à IS. Chaque niveau inclut un dictionnaire et une grammaire.

La génération

Elle s'opère en effectuant les étapes dans l'ordre inverse.

IS → ERS → ECS → EMS → phrase cible

Les grammaires et les dictionnaires sont communs à l'analyse et à la génération, les modules sont dits « réversibles ». Seuls les transferts diffèrent.

Le formalisme

En se fondant sur des mécanismes d'unification, EUROTRA tente de créer un nouveau formalisme, l'hypothèse de base étant que le processus de traduction peut être décrit comme une suite de projections, de telle sorte qu'un texte subit de l'entrée initiale à la sortie finale, un certain nombre de représentations grammaticales. Ces représentations sont déterminées par des critères linguistiques, qu'ils soient morphologiques, syntaxiques en surface ou en profondeur, et sémantiques.

La représentation IS

Elle est basée sur une grammaire de dépendance. La structure linguistique est analysée au moyen d'éléments noyaux (gov) et de dépendants, soient les arguments (arg) soient les modifieurs (mod).

Prenons une phrase s1 en allemand et sa traduction t1 en français, différentes dans la forme de surface :

s1 : Die Kommission verliert die Angst vor dem Rat.
t1 : La Commission cesse d'avoir peur du conseiller.

Prenons les structures IS de ces phrases s1 et t1 :

Le transfert s'opère à partir de cette représentation. En examinant les deux représentations, on constate qu'elles sont **isomorphes**, c'est-à-dire que les arbres ont la même géométrie et ne diffèrent que par les traits qui introduisent les valeurs lexicales. Quelques règles lexicales et un certain nombre de données abstraites portées sur les nœuds de l'arbre rendent possible un transfert simple alors que les phrases sont structurellement différentes et permettent de générer la phrase cible.

Le transfert

Le choix d'un modèle à transfert plutôt qu'un modèle à langage pivot s'explique par le fait qu'un grand nombre de langues sont traitées ici. Il implique pour chaque langue l'écriture d'un module de transfert et d'un module de génération, étant donné que la représentation du texte source contient des informations sur la structure et le vocabulaire de ce dernier, ce qui n'est pas le cas d'un modèle à langage pivot qui lui est censé représenter le sens profond du texte source et procéder par concepts. Les langues traitées ici étant apparentées, des structures de la langue source peuvent trouver des équivalents dans la langue cible. Le degré d'abstraction de la représentation IS et son caractère sémantique sont les garants d'une traduction de qualité.

Le logiciel

Le logiciel appelé **ETS** (Eurotra Translation System, version 2.0.1.) fonctionne sur un ensemble de plates-formes **UNIX**™, il est écrit en langages **Prolog** et **C** et est construit autour d'un nombre de progiciels commerciaux (compilateurs Prolog, système de gestion de base de données, analyseur grammatical SGML, etc.). Le système complet, incluant le matériel linguistique, occupe une mémoire disque de 100 Mo. La configuration type est une machine UNIX™ d'une vitesse d'au moins 3 Mips CPU et dotée d'une mémoire centrale d'au moins 8 Mo.

Il est écrit par un groupe qui appartient à la Commission et il est distribué aux équipes nationales qui effectuent le travail linguistique.

EUROTRA a choisi de créer un langage spécifique à ses besoins, écrit en Prolog. Quelques notions comme l'unification et le *backtracking* sont typiques de Prolog mais le formalisme d'EUROTRA est conçu pour permettre l'écriture de règles de grammaire de manière plus simple qu'en Prolog.

Les composantes majeures du système sont les suivantes :

- un compilateur de règles;
- un interpréteur de règles;
- une base de données;
- un *text handling*;
- une interface homme-machine.

Le compilateur de règles

Il traduit les règles de grammaire écrites dans le formalisme propre à EUROTRA sous forme de clauses Prolog. L'avantage principal résultant du compilateur de règles est la vérification de la syntaxe. Un certain nombre de règles (dites *lexicon rules*) peuvent être intégrées à une base de données lexicales et ne sont plus dans les fichiers générateurs.

Le compilateur de règles est un programme écrit en C. Il utilise un analyseur lexical généré par **lex** et un analyseur grammatical généré par **yacc** (**lex** et **yacc** sont des outils d'UNIX™).

L'interpréteur de règles

L'interpréteur de règles réalise le transfert et la consolidation d'un objet d'un niveau à un autre en appliquant les règles de grammaire compilées qui alimentent ces niveaux.

Le transfert est réalisé par le transducteur, la consolidation par le générateur.

Pour produire la représentation d'un objet à un niveau donné, à partir de sa représentation au niveau précédent, le transducteur prend chaque nœud de l'arbre de départ, en partant de la racine, lui applique les règles de transfert puis le générateur consolide et affine ce nœud, avant de passer au nœud suivant. Chaque nœud sera évidemment décomposé à son tour, s'il n'est pas une feuille de l'arbre et chacune de ses parties sera transférée et consolidée.

L'interpréteur de règles fait partie du cœur du logiciel Prolog.

La base de données

La base de données **EDB** (Eurotra Data Base) est utilisée pour stocker les règles lexicales qui composent les dictionnaires monolingues du système. Pour le moment, la composante base de données est toujours en cours de développement. Elle permettra dans le futur de stocker les règles lexicales des dictionnaires de transfert.

Une base de données est nécessaire pour un certain nombre de raisons :

- sans une base de données efficace, une plate-forme système raisonnable ne pourrait pas maintenir plus de 5000 entrées lexicales; le but d'EUROTRA est d'avoir un lexique d'environ 20 000 entrées par langue à la fin de 1992;
- l'administration du système est plus facile si les règles sont dans une base de données. Elle peut contenir, en plus des règles elles-mêmes, d'importantes informations annexes comme l'auteur des règles, la date de création, le numéro de version du système au moment de création de la règle, un commentaire, etc.;
- dans un environnement multi-utilisateurs, les fichiers sont peu pratiques, ils ne peuvent être mis à jour que par une seule personne à la fois; une base de données permet de mieux contrôler les accès multiples aux données;
- il est aussi plus efficace de ne compiler que les nouvelles règles entrées dans la base plutôt que le fichier entier.

L'implémentation actuelle est basée sur **UNIFY** ™, un système de base de données relationnelle, commercialisé par *Unify Corporation*, qui accepte le langage standard industriel de base de données **SQL**.

Le *Text Handling*

Le sous-système *Text Handling* est une interface entre le texte écrit et les objets que le système EUROTRA utilise de façon interne.

C'est un processus à deux étapes. À l'entrée du système au début de la phase d'analyse du texte, le processus est appelé segmentation. La première étape produit une forme canonique à partir du texte d'entrée avec les codes de traitement de texte remplacés par les codes de SGML. À cette étape les caractères nationaux (et autres caractères spéciaux) sont normalisés. La deuxième étape reconnaît les éléments tels que marques de ponctuation, nombres, codes, etc. et extrait les mots du texte générant un objet qui est transmis au système de traduction. À la sortie du système (à la fin de la phase de génération de texte dans le langage cible) le processus est appelé désegmentation, les deux mêmes étapes sont appliquées dans l'ordre inverse.

Le sous-système *Text Handling* est basé sur **SGML** (Standard Generalized Markup Language). SGML est une norme de l'**ISO** (International Standards Organisation).

L'interface homme-machine

C'est une interface entre l'utilisateur et le cœur du logiciel. Il y a actuellement deux interfaces dans le système, une interface pilotée par menus et une interface de commande. La première donne accès à toutes les fonctions principales du système. La seconde donne un contrôle supplémentaire à l'expert (le linguiste) qui veut utiliser les fonctions interactives de debogage.

L'interface homme-machine comprend un certain nombre de composantes :

- le système de menus,
- l'interpréteur de commandes,
- le système d'aide,
- les fichiers de paramètres.

Le système de menus

Le système de menus permet un accès facile au système. En utilisant les menus, un utilisateur promène ses objets linguistiques à travers les différents niveaux du modèle formel d'EUROTRA.

Les objets sont stockés dans des fichiers.

À chaque niveau, les objets résultats peuvent être vérifiés en imprimant les fichiers ou en utilisant le *pretty printer* qui met l'objet sous forme d'arborescence.

L'utilisateur peut envoyer des traitements en arrière-plan à partir du système de menus, cela lui permet de continuer à travailler (édition, visualisation, etc.) pendant que le travail en arrière-plan s'exécute. Le travail en arrière-plan est choisi parmi une liste de travaux qui ont été préalablement définis dans le fichier de paramètres de l'utilisateur. Le système de menus permet un contrôle absolu sur ces travaux, ils peuvent

être surveillés ou interrompus. Un fichier-journal est créé par le système et peut être visualisé à partir du système de menus.

Le système de menus est un programme écrit en C, qui utilise le module **curses**. **Curses** est une bibliothèque d'UNIX™ qui permet la gestion du fenêtrage de l'écran.

L'interpréteur de commandes

L'interpréteur de commandes est disponible à partir du système de menus. Il peut être également utilisé tout seul. Toutes les fonctions du système de menus sont disponibles à partir de l'interpréteur de commandes.

Un nombre de fonctions supplémentaires sont utilisables à partir de l'interpréteur de commandes :

- contrôle des objets en mémoire. Les objets peuvent être laissés en mémoire, au contraire du système de menus qui écrit les objets dans des fichiers;
- possibilité de suivre l'exécution de la traduction en regardant quelles règles sont appliquées et dans quel ordre; contrairement au système de menus qui permet de suivre l'exécution globale de la traduction, on peut demander à ne voir que l'exécution de certaines règles.

L'interpréteur de commandes est intégré dans le cœur du programme Prolog. Le programme Prolog analyse les commandes et les exécute directement.

Le système d'aide

Le système d'aide en ligne est disponible à partir du système de menus.

Il offre à l'utilisateur un ensemble de sujets qui sont couverts par le système. En choisissant un sujet, l'utilisateur peut lire la documentation sur ce thème.

Le système d'aide n'est pas disponible à partir de l'interpréteur de commandes.

Les fichiers de paramètres

Il y a deux fichiers de paramètres utilisés dans le système EUROTRA : le fichier de paramètres de l'utilisateur et le fichier de paramètres du système.

Le fichier de paramètres de l'utilisateur sert à mémoriser toutes les informations concernant l'environnement de travail. Par exemple, les fichiers et répertoires de travail, le nom de l'éditeur que l'utilisateur préfère utiliser, etc. L'utilisateur peut changer ces valeurs.

Le fichier de paramètres du système est mis à jour par l'administrateur système. Ce fichier contient des informations qui en général changent seulement quand une nouvelle version du logiciel est disponible. Il contient, entre autres, les chemins d'accès aux utilitaires servant au logiciel.

Les fichiers de paramètres sont des fichiers texte. Chaque ligne commence par une lettre qui définit le type du paramètre, suivi par le nom du paramètre lui-même, suivi par sa valeur.

Extrait d'un fichier de paramètre du système :

VE, ets201
P, PROLOG, /usr/eurotra/rel201/bin/et-prolog
P, PR-PAR, « -h 1500 -g 512 -l 512 -a 256 -t 64 »

La première ligne définit la version du logiciel. La seconde donne le chemin d'accès à l'interpréteur Prolog et la troisième donne les paramètres avec lesquels l'interpréteur Prolog doit être appelé.

Les perspectives

Nous sommes actuellement dans une phase de deux ans dite de transition du programme EUROTRA qui a pour but une réflexion sur l'état de l'art et sur les développements futurs du produit obtenu.

En matière de recherche communautaire, bon nombre de problèmes linguistiques n'ont pas été résolus, voire même abordés (problèmes épineux de la traduction du temps et de l'aspect, du traitement des composés des langues germaniques dans les dictionnaires, de l'anaphore...); il nous semble donc que ces axes de recherche peuvent être maintenus, et cela en coopération avec de futurs utilisateurs, lexicographes, traducteurs.

Puisque le thème de ces journées est *l'environnement du traducteur en l'an 2001*, nous aimerions revenir, en guise de conclusion, sur certains aspects techniques du logiciel et les possibilités qu'il peut offrir à un traducteur.

En effet, après la période d'euphorie des années soixante où l'on pensait que la machine traduisante pourrait tout faire, un certain nombre de recherches dans ce domaine ont été abandonnées.

Par la suite, des travaux menés sur des applications ont amené les chercheurs à voir les choses de manière plus modeste et à concevoir non plus la traduction comme automatique mais comme un outil d'aide au traducteur, efficace et souple. Nous pensons que l'avenir d'un programme tel qu'EUROTRA se trouve dans le développement de l'outil dont nous disposerons à la fin de 1992, en mettant l'accent sur sa souplesse et sa grande maniabilité y compris pour un novice; parmi les outils à développer, généralisables à d'autres applications de traduction automatique ou de traitement du langage naturel on citera :

- l'outil *text handling*,
- l'outil grammaire,
- l'outil lexique,
- l'outil traitement linguistique.

L'outil *Text handling*

Il permettra de manipuler des textes (sections, paragraphes, tableaux, notes de bas de page), le stockage de textes et outils documentaires permettant de produire des concordances, des fréquences, l'analyse orthographique et la segmentation de texte.

L'outil grammaire

Il consistera en plusieurs outils pour ceux qui écrivent les grammaires : codage de la grammaire, possibilité pour l'utilisateur d'écrire et de maintenir les grammaires dans le formalisme d'EUROTRA, et cela grâce à des éditeurs spécialisés et des menus; possibilité de définir des sous-grammaires pour des tests spéciaux ou des applications particulières.

L'outil lexique

Il comprendrait plusieurs outils pour des lexicographes et des linguistes : codage du lexique, possibilité d'éditer des entrées lexicales, le codage étant assisté par un menu, dans le même esprit que l'actuelle base de données; importation de lexiques; possibilité de définir des sous-lexiques; vérification de la cohérence du lexique.

L'outil traitement linguistique

Il consistera en plusieurs sous-modules qui aideront l'utilisateur durant les phases d'essais et de débogages :

- la machine virtuelle fournira les capacités d'un moteur d'inférences (unification, résolution de contraintes). Les opérations exécutées sont l'analyse grammaticale, l'évaluation des contraintes et la génération;
- un débogueur offrira plusieurs possibilités pour superviser les opérations de la machine virtuelle;
- la possibilité d'examiner des sous-résultats (texte ou graphique);
- un agent de transfert permettra à l'utilisateur de spécifier une application de traduction automatique basée sur le transfert.

La multiplicité des langues traitées par le système (72 paires) a imposé une stratégie de transfert simple à partir d'une représentation IS relativement abstraite en regard de la structure de départ. Ce passage obligé est la condition pour une traduction de qualité satisfaisante. De nombreuses et longues recherches sont encore nécessaires pour mettre au point cette représentation car les études menées de manière contrastive montrent que chaque fois qu'une information linguistique se réalise différemment d'une langue à l'autre, une représentation abstraite doit être mise au point, cohérente avec tous les phénomènes qui ont été calculés auparavant. Ceci demande un vaste programme de

recherche pour faire aboutir l'état des connaissances sur les neuf langues. À court terme on peut espérer que le système produira des traductions laissant à l'humain le choix de la traduction correcte, à long terme nous souhaitons que le programme EUROTRA aura montré que les recherches sur l'interface sémantique et sa portée interlinguistique doivent être poursuivies.

Bibliographie

BECH A. (1990) « The Virtual Machine », *Eurotra Reference Manuel 6.1*, Technical Report of Eurotra, Luxembourg.

DANLOS L., BECH A., CAROLI F., DAILLE B., MESLI N., NAMER F., NOHR S. (1989) « Support Verbs and Predicative Nouns in a Machine Translation System », *Eurotra Reference Manuel 6.0*, Technical Report of Eurotra, Luxembourg.

DANLOS L., LAURENS O. (1991) « Présentation du projet Eurotra et des grammaires d'Eurotra-France », *Rapport technique Eurotra-Talana*, n° 1, Paris.

DANLOS L., NAMER F. (1991) « Les constructions à verbe support : propriétés linguistiques et implémentation dans Eurotra », *Rapport technique Eurotra-Talana*, n° 4, Paris.

EUROTRA (1991) « Eurotra Reference Manual 7.0 », *Technical Report of Eurotra*, Luxembourg.

GAJEK O., LUKS K., MEYER R., NETTER K., SCHUETZ J., THURMAIR G., WEHRMEYER A. (1990) « Requirements Study for a EUROTRA-II Software Environment », *Internal Progress Report*, Luxembourg.

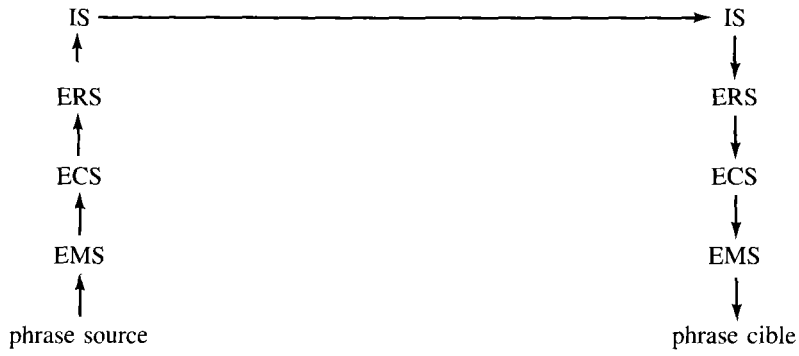
GAJEK O., LUKS K., MEYER R., SCHUETZ J., THURMAIR G., WEHRMEYER A. (1990) « Assessment Study for a EUROTRA-II Software Environment », *Internal Progress Report*, Luxembourg.

NAMER F. *et al.* (1990) *Rapport d'implementation Eurotra-France*, Paris.

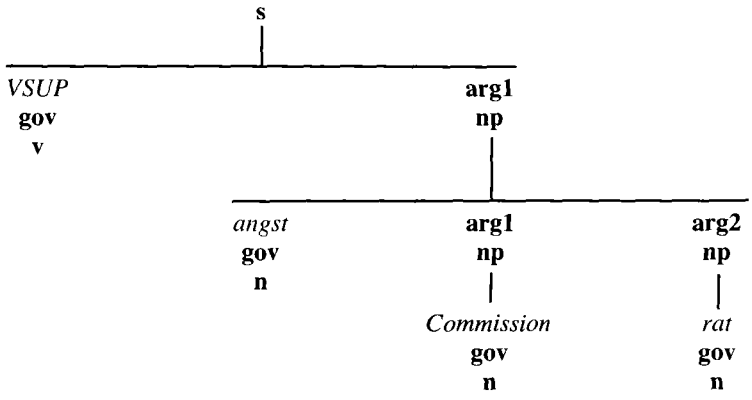
SCHUETZ J. (1990) « A Multi-Agent Architecture for a NLP Development Environment », Ms., IAI, Saarbrücken.

SCHUETZ J. (1990) « An Architecture for Reusable Lexical Resources in a Multi-theoretical Environment », Ms., IAI, Saarbrücken.

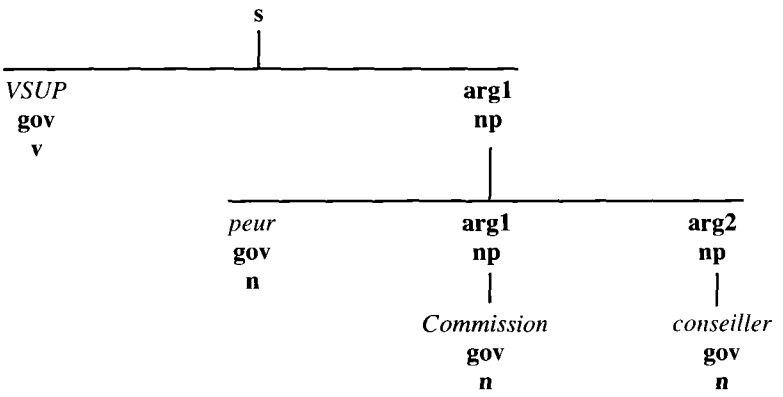
Annexes


IS
IS
ERS
ERS
ECS
ECS
EMS
EMS
phrase source
phrase cible

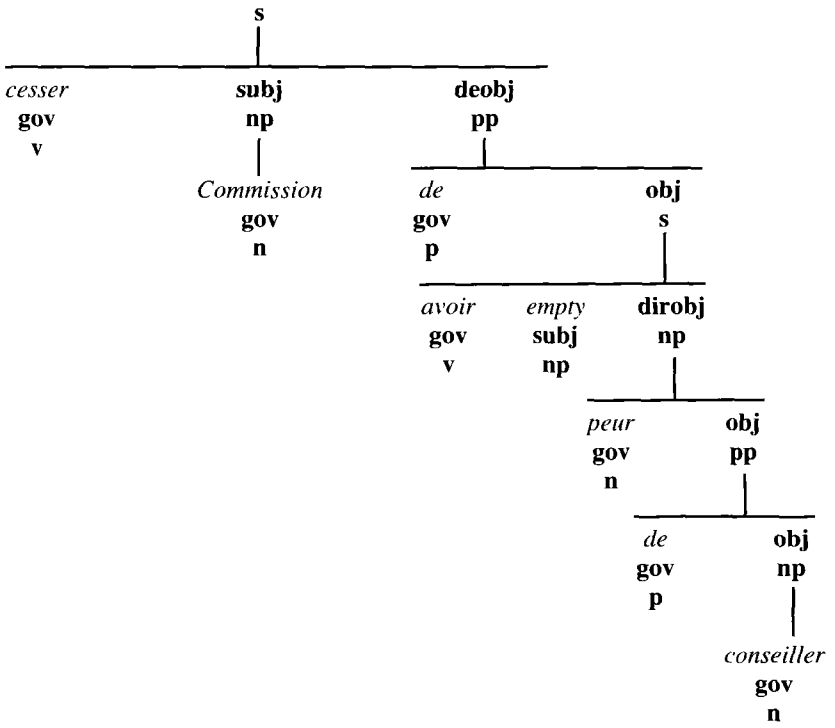
ISD


s
VSUP
gov
v
arg1
np
angst
gov
n
arg1
np
Commission
gov
n
arg2
np
rat
gov
n

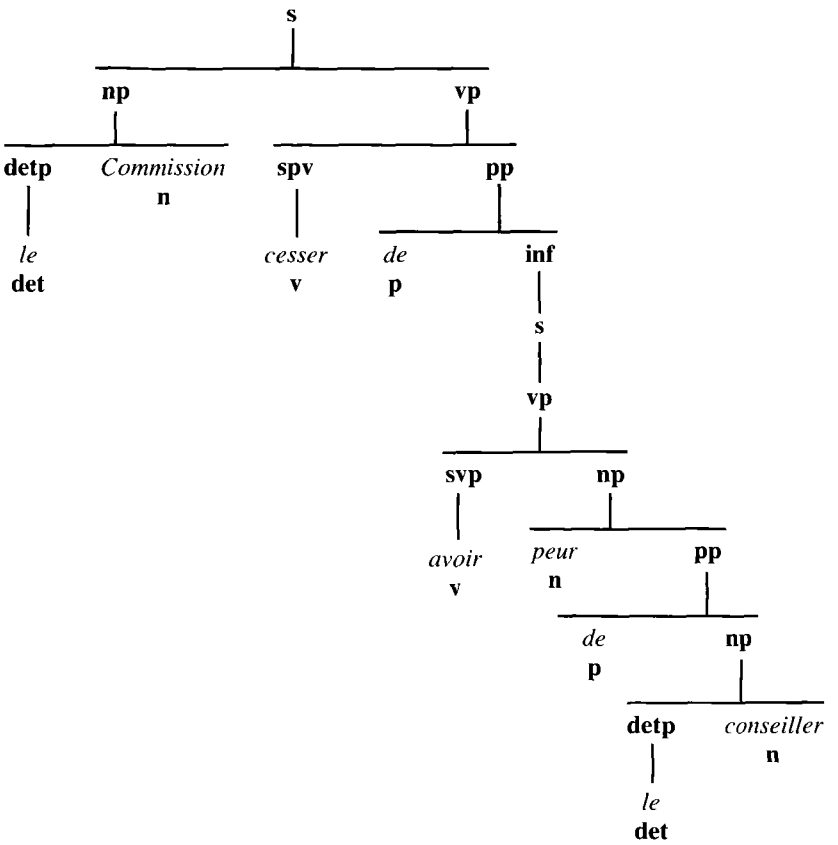
ISFR


s
VSUP
gov
v
arg1
np
peur
gov
n
arg1
np
Commission
gov
n
arg2
np
conseiller
gov
n

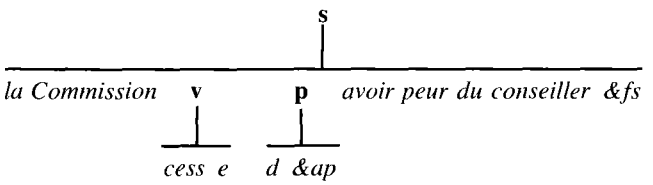
ERSFR


s
cesser
gov
v
subj
np
Commission
gov
n
deobj
pp
de
gov
p
obj
s
avoir
gov
v
empty
subj
np
dirobj
np
peur
gov
n
obj
pp
de
gov
p
obj
np
conseiller
gov
n

ECSFR



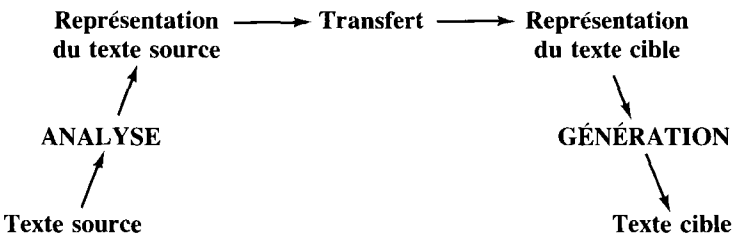
EMSFR

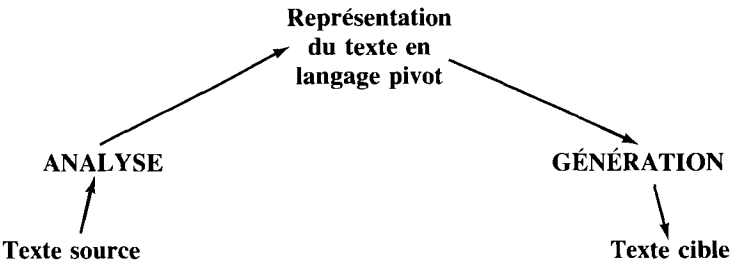


TEXTE

%% LANGID = fr
%% 1

la Commission cess e d avoir peur du conseiller.



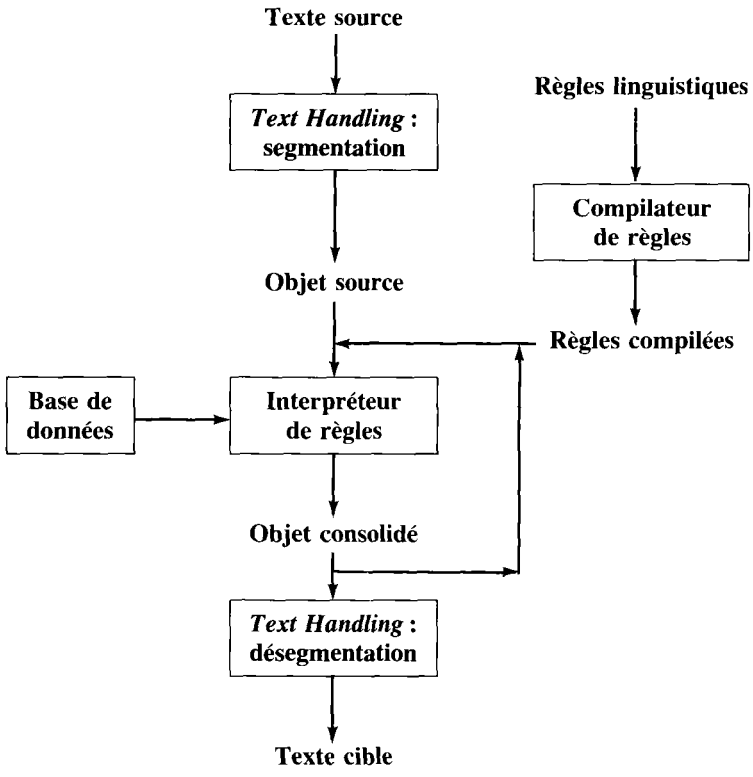


Exemple de phrase à traduire :

Heureusement, la Commission prend des mesures importantes.

Objet utilisable par l'interpréteur de règles, après segmentation :

```
:filetype:object.
language(text,fr).
obj(text,fb(cat,io,lex,rend),l:[]:{cat=s,io='0'}.[
l:[]:{lex='Heureusement',rend=no}.[]
,l:[]:{lex='&cm'}.[]
,l:[]:{lex='la',rend=no}.[]
,l:[]:{lex='Commission',rend=no}.[]
,l:[]:{lex='prend',rend=no}.[]
,l:[]:{lex='des',rend=no}.[]
,l:[]:{lex='mesures',rend=no}[]
,l:[]:{lex='importantes',rend=no}.[]
],label([unknownfile/(text/1),'phrase'])).
```



Eurotra Translation System ets v2.1.0.9 26 Feb 1991

Dataset : Batch Pid : Dir :

Ets Toolbox

0. Unify User Environment
1. Segment Input Text
2. Compile Rules
3. Translate Objects
4. Invoke Command Interpreter
5. Edit File
6. Review Paramaters
7. Object Handling
8. Batch Process Control
9. Document Processing

Enter your choice :

Help ^ Xit !Shell ^ Wash

9

ODILE 2 un outil pour traducteurs occasionnels sur Macintosh

Mathieu LAFOURCADE

GETA, Institut IMAG (UJF & CNRS), Grenoble, France

Résumé

La conception d'outils informatisés d'aide aux traducteurs occasionnels est délicate. En effet, il faut intégrer des outils déjà existants et interagir avec les applications du commerce. Nous avons réalisé une application sur Macintosh, Odile, qui permet, en utilisant un lemmatiseur, de fabriquer à partir d'une chaîne sélectionnée un ensemble de clés de recherche, de les compléter si nécessaire « à la main », et de les utiliser pour accéder à des dictionnaires usuels supportés par un utilitaire du commerce. L'architecture d'Odile est modulaire. Elle repose sur la définition d'interfaces génériques avec des lemmatiseurs et des outils dictionnaires. Après avoir présenté la version actuelle d'Odile, on indique comment l'améliorer. En particulier, l'introduction d'applications communicantes au niveau du système d'exploitation devrait permettre la réalisation de fenêtres « vocables », et l'élimination presque totale des interventions de l'utilisateur.

Introduction

Dans le cadre de la THAM (Traduction humaine assistée par la machine), nous nous intéressons à la création d'outils destinés aux traducteurs occasionnels. Il ne s'agit donc pas d'outils pour professionnels, qui peuvent être dédiés, mais de possibilités supplémentaires offertes aux utilisateurs de traitements de texte, tableurs, SGBD et autres outils du commerce. C'est pourquoi nous n'envisageons pas pour l'instant

d'accès à des « bases de bitextes » résultant de traductions antérieures (comme TSS d'ALPS sur IBM [8]).

Ce genre d'outil est destiné à un public non spécialisé (ni informaticien, ni linguiste), et vise à automatiser le processus de traduction. Si l'on estime qu'environ trente pour cent du temps de traduction est passé à chercher dans des dictionnaires, grammaires et autres, il semble clair qu'une informatisation de ce genre de tâches augmente la productivité. Les dictionnaires de tels outils doivent être, au moins partiellement, modifiables par l'utilisateur. Il faut aussi éviter de forcer l'utilisateur à taper une « forme canonique » (« cheval » pour « chevaux »). Pour l'instant, seuls des outils dédiés comme SISKEP [14] ont ces deux possibilités. Il est donc intéressant de les offrir à l'intérieur d'un outil ouvert. D'autre part, ce type d'outil peut servir à d'autres choses qu'à la traduction, les dictionnaires pouvant aussi bien être des fiches techniques ou des thesaurus.

La première partie présente de manière plus précise les motivations, les désirs et les contraintes que nous avons pris en compte pour la spécification d'Odile. La partie centrale de l'exposé est consacrée à la description de la version 2 d'Odile, une application sur Macintosh qui intègre un lemmatiseur et un outil dictionnaire. Odile permet, en utilisant un lemmatiseur, de fabriquer à partir d'une chaîne sélectionnée un ensemble de clés de recherche, de les compléter si nécessaire « à la main », et de les utiliser pour accéder à des dictionnaires usuels supportés par un utilitaire du commerce, WinTool™. Sa réalisation a été l'occasion de définir une interface universelle pour les lemmatiseurs. L'essentiel de la première version a été réalisé dans le cadre du mémoire CNAM d'Isabelle Tomasino [13]. La version 2 correspond à une nouvelle spécification de l'interface homme-machine, de l'interface logicielle et des fonctionnalités. La dernière partie montre un certain nombre d'améliorations possibles. En particulier, l'introduction d'applications communicantes au niveau du système d'exploitation devrait permettre la réalisation de fenêtres « vocables », et l'élimination presque totale des interventions de l'utilisateur.

Spécifications

Motivations

Les problèmes avec les outils existants

L'inconvénient majeur des outils existants est qu'ils sont souvent fermés (cas de WordFinder sur Macintosh). Le traducteur occasionnel ne peut donc en modifier les dictionnaires ni les règles. L'utilisateur reste donc dépendant du contenu des bases fournies avec de tels outils.

D'autres outils sont largement dépendants d'un environnement proposé en même temps (SISKEP de l'USM [14] et [15] et TAIM extension de Palantir de Melby [9]). L'utilisateur est alors obligé d'apprendre l'utilisation d'un nouvel éditeur.

Des outils ouverts existent (Mercury [10] , WinTool [16]) mais imposent d'entrer une clé de recherche correcte. L'utilisateur ne peut pas rentrer des clés approximatives.

On constate donc qu'aucun des outils proposés n'est à la fois ouvert et indépendant d'un environnement. De plus, ils obligent l'utilisateur à taper une « forme canonique » pour effectuer la recherche.

Les clés de recherche doivent être générées automatiquement ou à la main

Pour le français, il est souhaitable d'avoir un système automatique qui élabore les clés correspondant à un morceau de texte. De plus, l'utilisateur doit pouvoir gérer des clés « personnelles », car un dictionnaire automatique ne sera jamais ni complet ni parfait. Il est nécessaire de pouvoir introduire des termes composés, par exemple pour trouver « manette de sécurité » à partir de « manette de » sans se retrouver sur « mangeoire » après tous les « manette de... ». Il est également pratique de spécifier des équivalents de sigles, par exemple qu'OTAN correspond à NATO et de ne pas avoir à taper « Organisation du Traité de l'Atlantique Nord ».

Un générateur automatique de clés doit rester simple. Par conséquent, on ne le munira pas de fonctions complexes permettant, par exemple, de retrouver « mettre les points sur les i » à partir de « dans les copies de ses élèves, elle ne mettait pas toujours et partout les points sur les i ».

Un champ d'expérimentation avec une interface conviviale

La manipulation classique dans les outils d'aide à la traduction est de laisser l'utilisateur sélectionner une clé et de la chercher dans un dictionnaire. Nous pensons que ce type d'interaction peut être amélioré.

Désirs

Accessibilité universelle (programme résident)

L'utilisateur veut pouvoir utiliser les logiciels dont il a l'habitude en même temps que des outils d'aide à la traduction. On désire donc un outil indépendant de toute application particulière (voir Mercury [10]).

Généricité

Nous voulons pouvoir adapter n'importe quel lemmatiseur ou outil dictionnaire. PILAF [4] est adapté aux langues romanes. D'autres lemmatiseurs seraient adaptés à d'autres types de langues. On pourrait également inclure des lemmatiseurs « à dictionnaires restreints » (Palmer [12]), munis de méthodes heuristiques pour le traitement des mots inconnus.

En tant que chercheurs, nous voyons aussi là un moyen de rendre accessible les produits de laboratoire à un grand nombre d'utilisateurs. En effet, développer un environnement convivial et fiable, ou bien adapter les données d'un outil à un autre, représente un travail de développement trop lourd pour des équipes de recherche.

Convivialité

L'utilisateur veut lancer un processus de recherche et ne plus s'occuper de rien. L'interactivité doit donc être faible. Cela peut se résumer par la formule suivante : « give and leave ». D'autre part, l'utilisateur doit en regardant avoir l'intuition de ce qu'il peut faire (« look and feel »). Un certain nombre de principes d'ergonomie doivent être respectés (Apple [1] et [2] et Coutaz [5]).

Enfin, la rapidité d'exécution est primordiale : sans elle, l'utilisateur préférera ses dictionnaires papier.

Contraintes

Nous avons choisi de réaliser Odile sur Macintosh, en intégrant un lemmatiseur tiré du système PILAF [3] et un outil dictionnaire disponible (WinTool™).

Le Macintosh

Cette gamme de micro-ordinateurs offre une interface conviviale. L'interface générale d'Odile suit les spécifications préconisées par Apple.

Pilaf

Pour des raisons de rapidité d'implantation, nous avons adapté un lemmatiseur déjà existant sur le Macintosh (Courtin *et al.* [4]). Le moteur de PILAF existait en langage Pascal sur micro-ordinateur (compatible PC). Un premier portage sur le Macintosh a été fait dans le cadre du PRC-CHM en 1988-89 avec un travail d'adaptation sur le lemmatiseur. Un dictionnaire d'environ 25 000 lemmes (produisant environ 175 000 formes de sortie) provenant du travail de l'équipe TRILAN. Il s'agit d'une adaptation du dictionnaire DELAS du LADL (M. Gross).

WinTool™

WinTool™ est un outil dictionnaire déjà existant dans le commerce. Pour l'intégration, nous avons eu accès au code. D'autre part, WinTool™ est très rapide, peu cher et ne prend pas beaucoup de place mémoire. En effet, la taille du code est d'environ 100 Ko et la taille des fichiers représente environ 50 octets par articles.

Odile version 2

Odile version 2 est une réécriture complète de l'interface et une redéfinition des fonctionnalités de la première version d'Odile [13]. Des démonstrations sont possibles, soit du français vers l'arabe (avec système arabe, coopération avec l'IERA dans le cadre RFIL), soit du français vers le malais.

Interface et fonctionnalités

Exemple de session

Comment se présente Odile? Plutôt que de décrire les unes après les autres les fonctionnalités de cet outil, nous présentons le déroulement d'un scénario type. Supposons qu'un chercheur doive traduire son dernier article en vue d'une présentation à un congrès. Pour cela, il utilise son traitement de texte favori et traduit les paragraphes les uns après les autres directement sur le document. Cependant, il hésite pour traduire le fragment « mettent les pieds dans le plat » (« (they) spill the beans »).

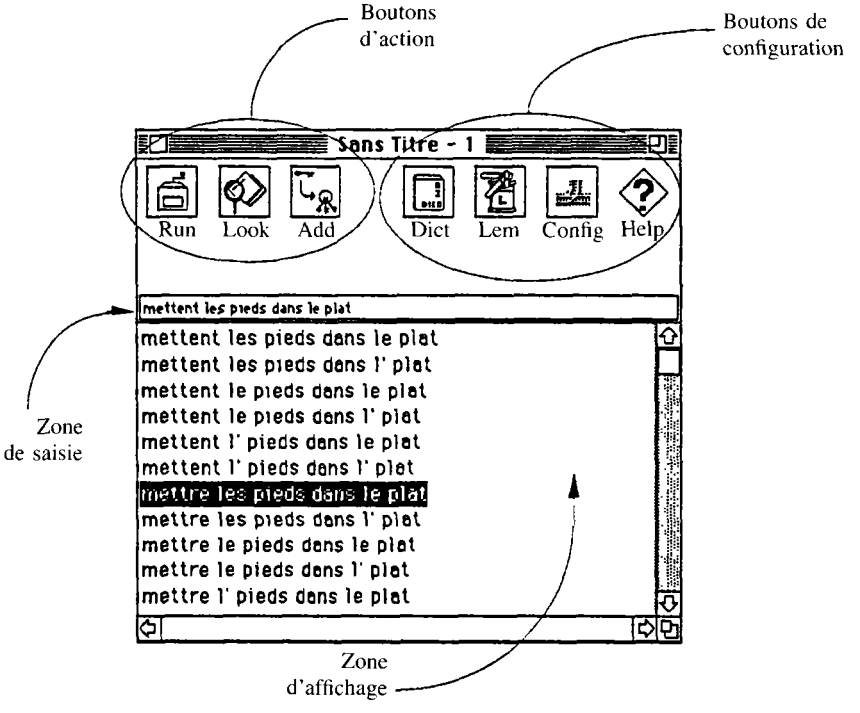


Figure 1 : La fenêtre principale d'Odile. Cette fenêtre est divisée en deux parties : son bord (partie supérieure) et son contenu (partie inférieure). Le bord comporte deux séries de boutons. La première correspond à des boutons performant des actions, la seconde permet de configurer les fonctionnalités d'Odile. Le contenu est une zone de texte libre ou sont affichées les clés obtenues par combinaison des formes et des lemmes.

Il sélectionne ce fragment, le copie et lance Odile. La fenêtre d'Odile s'affiche. L'utilisateur colle alors le fragment dans la *zone de saisie* d'Odile (voir figure 1) et lance la génération de clés avec le bouton *Run*.Les clés (potentielles) calculées s'affichent alors dans la *zone d'affichage*. On peut automatiser le processus en configurant Odile différemment. Les différentes possibilités sont illustrées plus bas avec l'automate de contrôle.

L'utilisateur sélectionne la clé désirée et lance la recherche (bouton *Look*) dans le dictionnaire. La recherche se fait par défaut dans le dictionnaire usuel, mais on peut aussi chercher dans d'autres dictionnaires.

Si les clés proposées ne lui conviennent pas, l'utilisateur a toujours la possibilité d'en entrer une directement au clavier. Il peut aussi ajouter une clé dans son « lexique personnel ». Il peut bien sûr spécifier le ou les lexiques personnels où se fera la recherche.

Une fois la recherche terminée, la fenêtre de l'outil dictionnaire (WinTool™) s'affiche. Si la clé a été trouvée, on y voit l'article correspondant (clé + contenu). Si la clé était absente du dictionnaire, la fenêtre contient l'article de clé immédiatement inférieure.

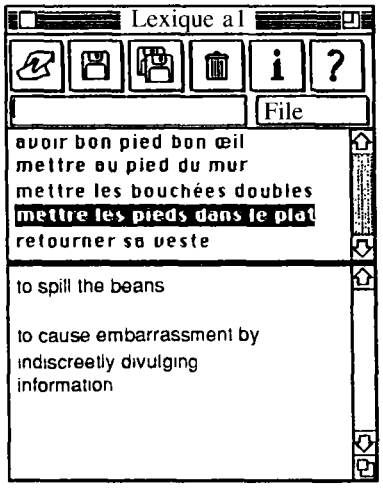


Figure 2 : Une fiche WinTool coorespondant à l'expression « mettre les pieds dans le plat ».

Les icônes de l'outil dictionnaire permettent la gestion de la base ainsi que la navigation. Une fois les informations trouvées, l'utilisateur retourne à son traitement de texte. Il peut garder la fenêtre affichée à côté de la fenêtre contenant son article de manière à avoir toujours accès à ce qu'il a cherché. Il peut également copier et coller les informations.

L'automate de contrôle

La séquence de traitements effectuée à partir de l'introduction d'un fragment dans la zone d'édition et de l'appui du bouton *Run* peut s'illustrer par un automate (figure 3).

Cet automate permet de définir le traitement à appliquer en fonction des options choisies par l'utilisateur. Pour le moment, deux options (majeures) sont définies :

- vérifier d'abord si la chaîne rentrée constitue une entrée de lexique ou de dictionnaire. Si oui, il n'y a pas de lemmatisation et Odile affiche directement l'article correspondant;
- utiliser un lexique personnel pour produire directement des clés de recherche.

Dans ce cas, les clés potentielles sont utilisées pour accéder à un lexique personnel (« trousseau »). Le contenu de ces clés est ensuite « interprété » comme une liste de clés, qui sont alors utilisées pour chercher dans le dictionnaire.

Dans l'automate, ces deux options correspondent respectivement aux conditions C1 et C2.

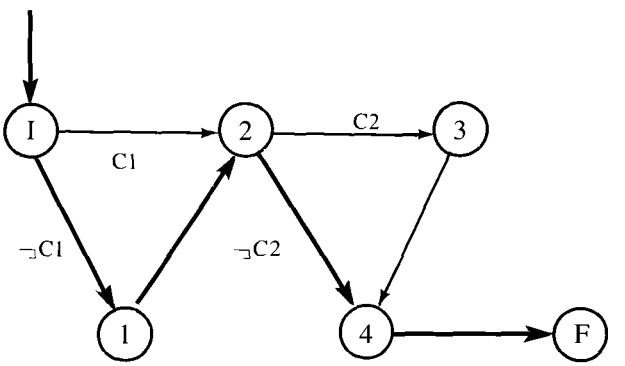


Figure 3 : L'automate de contrôle d'Odile.
Les états I et F sont respectivement les états initial et final. Les Ci sont les conditions pour passer d'un état à l'autre. Les arcs vides sont des transitions par lesquelles on passe sans conditions. Le chemin en gras correspond au plus court chemin possible, celui où l'utilisateur n'utilise aucune des options de traitement.

L'automate a six états. I est l'état initial, F l'état final;

- état 1 : on lemmatise F (la forme d'entrée). On obtient [CP]+, une liste non vide de clés potentielles CP;
- état 2 : état vide, on ne fait rien (c'est un état de transition);
- état 3 : on effectue la recherche des [CP]+ dans le(s) lexique(s) personnel(s) courant(s). [CP]+ devient alors la liste des clés trouvées comme contenu des entrées du(des) lexique(s) personnel(s);

– état 4 : on effectue une recherche dans la(les) base(s) dictionnaires des entrées [CP]+ et on affiche la première clé trouvée. On affiche également dans la fenêtre Odile la liste des clés potentielles.

Standardisation des lemmatiseurs (définition d'une interface universelle pour les lemmatiseurs)

Nous avons élaboré des points d'entrée standard et des structures de sortie génériques pour tous les lemmatiseurs possibles. Les structures de données d'entrée sont décrites avec les points d'entrée.

Structure de données d'entrée pour les lemmatiseurs

Une chaîne constitue les données d'entrée. Elle a la forme <chaîne sans espace> (espace <chaîne sans espace>)*. Une telle chaîne est dite normalisée.

Structure de données de sortie pour les lemmatiseurs

La structure de sortie est un graphe avant contenant les informations suivantes :

– les nœuds représentent des formes ou des lemmes;

– les arcs représentent les chemins possibles parmi ces nœuds;

– chaque nœud contient jusqu'à dix informations (entre parenthèses sont indiqués le nom de la variable et son type) :

1) le numéro du nœud (Numéro : entier);
2) son identification, obligatoire (valChaîne : chaîne) et ne contenant pas de blanc;
3) son type, obligatoire (TypeChaîne) : Forme ou Lemme (F | L);
4) sa nature, facultative (Nature) : Occurrence complète, Tournure, morceau Gauche, Interne, ou Droit (D | O | I | G | T);
5) sa localisation, obligatoire (Origine : entier), désigne l'origine (0 désignant la position précédant le premier caractère) de ValChaîne;
6) sa longueur, obligatoire (Longueur : entier), qui est le nombre de caractères couverts par le nœud (la longueur de la sous-chaîne traitée dans la chaîne d'entrée);
7) sa catégorie morphologique (Catégorie : chaîne), facultative (par exemple : (Catégorie 'adv) ou (Catégorie 'subc) ou encore (Catégorie 'dept));
8) ses variables morphologiques (Variables : liste), facultatives (par exemple : (Variables('plus 'mas 'tre 'suj)) ou (Variables((nombre 'plus) (genre 'mas) (fonction 'suj)));

9) des informations supplémentaires (Reste : liste), facultatives, qui pourraient être utilisées dans l'avenir (par exemple, relief typographique, domaine du document, etc.);

10) les numéros des nœuds successeurs du nœud courant dans le graphe (Suivant : liste d'entiers), la liste vide (()) dénotant l'absence de successeur.

Cette structure est représentée dans un fichier texte, qui contient la suite des nœuds dans l'ordre ascendant des numéros. Chacun des nœuds est codé (à la LISP) comme suit :

(((Numéro *Numéro*) (ValChaîne *ValChaîne*)
(TypeChaîne *TypeChaîne*) [(Nature *Nature*)]
(Origine *Origine*) (Longueur *Longueur*)
[(Catégorie *Catégorie*)] [(Variables ({*Variables*}+))]
[(Reste ({*Reste*}+))])
(Suivants ({*Suivant*}+)

Notation : les caractères spéciaux { } [ ] * + sont utilisés pour coder des expressions régulières. Les parties entre crochets ([ ]) sont facultatives, les symboles en italiques dénotent les valeurs et les accolades ({ }) permettent des regroupements. La construction {*x*}+ signifie que x peut être répété une ou plusieurs fois, et {*x*}* signifie que x peut être répété zéro ou plusieurs fois.

Points d'entrée pour les lemmatiseurs

Procédure Lem-ChargerDonnees;

La procédure Lem-ChargerDonnees charge les données nécessaires à l'analyse des occurrences fournies en entrée.
L'appel de cette procédure doit se faire une seule fois (avant de lancer la procédure Lem-Lemmatiser!).

Procédure Lem-Lemmatiser (Chaîne : Chaîne;
var StructSortieLem : Fichier texte);

La procédure Lem-Lemmatiser lemmatise le groupe d'occurrences *(Chaîne)* fourni en entrée. Elle construit la structure des résultats de lemmatisation *StructSortieLem*.

Chaîne est la chaîne de caractères à lemmatiser. Il s'agit d'une chaîne « normalisée », décrite par l'expression régulière : <chaîne sans blanc>(blanc <chaîne sans blanc>)*.

StructSortieLem est la structure de sortie, construite dans tous les cas.

Exemple

Avec un lemmatiseur approprié, l'analyse de la chaîne « les poules couvent » peut donner le fichier texte résultat suivant. Les informations optionnelles Catégorie et Variables ne figurent pas.

(((Numéro 1) (ValChaîne 'les) (TypeChaîne 'F) (Nature 'O) (Origine 1) (Longueur 3)) (Suivants (4 5)))

(((Numéro 2) (ValChaîne 'le) (TypeChaîne 'L) (Nature 'O) (Origine 1) (Longueur 3)) (Suivants (4 5)))

(((Numéro 3) (ValChaîne 'il) (TypeChaîne 'L) (Nature 'O) (Origine 1) (Longueur 3)) (Suivants (4 5)))

(((Numéro 4) (ValChaîne 'poules) (TypeChaîne 'F) (Nature 'O) (Origine 5) (Longueur 6)) (Suivants (6 7 8 9)))

(((Numéro 5) (ValChaîne 'poule) (TypeChaîne 'L) (Nature 'O) (Origine 5) (Longueur 6)) (Suivants (6 7 8 9)))

(((Numéro 6) (ValChaîne 'couvent) (TypeChaîne 'F) (Nature 'O) (Origine 11) (Longueur 7)) (Suivants ()))

(((Numéro 7) (ValChaîne 'couvent) (TypeChaîne 'L) (Nature 'O) (Origine 11) (Longueur 7)) (Suivants ()))

(((Numéro 8) (ValChaîne 'couver) (TypeChaîne 'L) (Nature 'O) (Origine 11) (Longueur 7)) (Suivants ()))

(((Numéro 9) (ValChaîne 'couver) (TypeChaîne 'L) (Nature 'O) (Origine 11) (Longueur 7)) (Suivants ()))

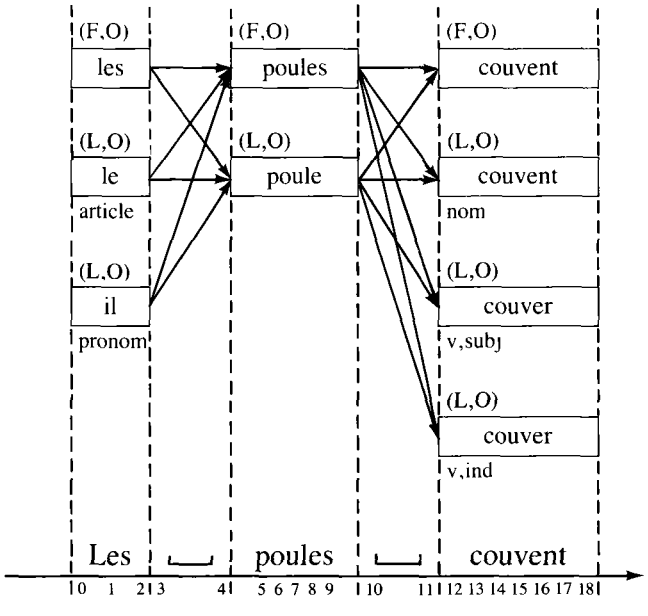


Figure 4: Résultat de la lemmatisation de la chaîne (normalisée) « les poules couvent ».

Intégration d'un outil dictionnaire

Pour qu'on puisse l'intégrer à Odile, un outil dictionnaire (et WinTool) doit avoir les points d'entrée suivants :

Chargement et déchargement de l'outil dictionnaire

Fonction LoadDT : Erreur;

Cette procédure permet de charger les données nécessaires pour le fonctionnement de l'outil dictionnaire

Fonction UnloadDT : Erreur;

Cette procédure doit être appelée quand on ne veut plus utiliser l'outil dictionnaire.

Contrôle de l'outil dictionnaire

Procédure ControlDT (Event : Événement);

On passe l'événement *(Event)* à l'outil dictionnaire.

Gestion des bases

Fonction OpenDTBase (BaseName : chaîne;
InfoNeeded : data;
Lexique : Booléen) : Erreur;

Cette fonction ouvre une base de l'outil dictionnaire.

Fonction CloseDTBase (BaseIndex : entier) : Erreur;

Cette fonction ferme une base de l'outil dictionnaire.

Gestion des articles

Fonction FindArticle (Key : string;
BaseIndex : integer;
Seuil : integer) : Erreur;

Cette procédure permet de rechercher des articles dont l'entrée est proche de *Key* dans la base *BaseIndex*. Proche signifie que la distance entre les articles trouvés et *Key* est inférieure à *Seuil*. Cette fonction écrit dans un fichier texte une liste de doublets. Chaque doublet représente la clé de l'article trouvé et la distance de cet article à *Key*. Le format du fichier est :

[(c,d)]* ()

c est une clé d'article et *d* est la distance de c par rapport à *Key*. Une liste vide (*()*) indique la fin de la liste.

On utilise cette procédure quant on veut juste s'assurer de l'existence d'un article dans une base, mais que l'on ne veut pas récupérer l'information associée.

Fonction GetArticle (Key : chaîne;
KeyNumber : integer;
BaseIndex : entier;
var Content : liste de données;
Visible : boolean) : Erreur;

Cette procédure permet d'obtenir les informations associées à l'article repéré par la clé *Key* de la base repérée par *BaseIndex*.

S'il existe plusieurs articles ayant la même clé, on obtiendra les *KeyNumber* premiers. On utilisera cette fonction pour récupérer les informations d'un article dont on sait qu'il existe dans la base.

Fonction PutArticle (Key : string;
Content : data;
BaseIndex : integer) : Erreur;

Cette procédure crée un nouvel article de dictionnaire dans la base référencée par *BaseIndex*. Cet article a pour clé *Key* et pour contenu les informations (Texte) référencées par *Content*.

Perspectives

Nous travaillons actuellement sur un certain nombre d'améliorations par rapport à la version actuelle.

Accélération du traitement

Le lemmatiseur d'origine (PILAF [4] et [3]) n'est pas vraiment adapté à ce que nous voulons faire. La structure de dictionnaire couplée au lemmatiseur n'est pas très efficace. Elle peut être améliorée, diminuant ainsi grandement le temps de génération de clés.

Nous cherchons également à adapter un ou plusieurs autres lemmatiseurs, puisqu'Odile a été conçu pour offrir une telle modularité.

Amélioration de l'interface utilisateur

L'idée d'une fenêtre « vocable » (d'après la revue du même nom) enfin possible avec le système 7.0 d'Apple [2] serait très intéressante. Une telle fenêtre serait disposée à côté d'une fenêtre de travail et contiendrait les articles des dictionnaires usuels correspondant aux termes présents dans la fenêtre. Elle serait effectivement liée à la fenêtre de travail en ce sens que l'outil irait « par avance » rechercher les informations associées au vocabulaire du texte affiché.

Conclusion

L'augmentation des performances des micro-ordinateurs permet maintenant la conception d'outils grand public en THAM et dans bien d'autres contextes. Les outils plus spécifiques s'intégreront à des environnements spécialisés pour traducteurs professionnels.

Une version d'Odile existe déjà. La version suivante, en préparation, sera une refonte complète tournant sous la version 7 du système d'exploitation du Macintosh, et organisée autour d'applications communicantes. Elle devrait permettre la réalisation de fenêtres « vocables », ce qui semble impossible ou extrêmement délicat avec les versions précédentes du système.

En effet, la réalisation technique d'outils dépend beaucoup des possibilités de l'environnement en général, et du système d'exploitation en particulier. Certains systèmes d'avant-garde offrent de bonnes idées. Par exemple, la Communication Inter-Application de la version 7.0 du Macintosh doit permetre un prototypage poussé des fenêtres vocables en attendant la généralisation de ce concept à d'autres environnements. L'émergence de standards ou de systèmes d'exploitation équivalents comme UNIX avec XWindow/Motif et OS2/PM en permettra sans doute la large diffusion.

La démarche adoptée consiste à développer les idées nouvelles sur le premier système disponible permettant leur réalisation, tout en prévoyant une adaptation ultérieure pour les autres environnements fonctionnellement équivalents qui ne manqueront pas de voir le jour.

Bibliographie

[1] APPLE Computer Inc (1985) *Inside Macintosh*, Volume I, Addison-Wesley Publishing Company, Inc., 550 p.

[2] APPLE Computer Inc (1990) *Inside Macintosh*, Volume V, Addison-Wesley Publishing Company, Inc., 1200 p.

[3] CHAPON L., DJAMEI M., DJOHARIAN P., MORENO F. (1989) *Étude et révision du système PILAF, Portage vers le Macintosh, Réalisation d'un lemmatiseur, Rapport final de projet,* Projet DESS-IDC, juin 1989.

[4] COURTIN J., DUJARDIN D., GENTHIAL D., KOWARSKI I., COHARD B., STRUBE DE LIMA V. « Le système PILAF », document interne LGI (IMAG) présenté au *PRC Communication Homme-Machine*, février 1988.

[5] COUTAZ J. (1988) *Interface Homme-Ordinateur : conception et réalisation*, thèse de doctorat d'État (UJF), décembre 1988, 402 p.

[6] HUNDT M. G. (1982) *Working with the Weidner Machine-Aided Translation System, Practical Experience of Machine Translation*, V. Lawson (éd.), North-Holland Publishing Compagny, ASLIB, p. 45-51.

[7] KJAERSGAARD P. S. (1987) « REFTEX — Un progiciel pour la traduction assistée par ordinateur », *Computer and Translation*, 7 p.

[8] LONSDALE D. (1984) « Notes on Interactive Translation », *Automated Language Processing Systems (ALPS)*, Provo, Utah, 10 p.

[9] MELBY Alan K. (1978) « Design and Implementation of a Computer-Assisted Translation System », Présenté à la *7th International Conference on Computational Linguistics (COLING-78)*, Bergen, Norvège, 14-18 août 1978, 28 p.

[10] MELBY Alan K. (1984) *MERCURY™ Card File and Glossary Manager, User's Guide,* LinguaTech, Provo, 114 p.

[11] META, (1990) *Journal des traducteurs*, 35e anniversaire, Les Presses Universitaires de Montréal, Vol. 35, n° 1, mars 1990.

[12] PALMER P. (1990) *Étude d'un analyseur de surface de la langue naturelle, Application à l'indexation automatique de textes,* thèse (UJF), septembre 1990, 292 p.

[13] TOMASINO I. (1990) *ODILE : un outil d'intégration extensible de dictionnaires et de lemmatiseurs*, mémoire (CNAM), décembre 1990, 151 p.

[14] TONG L. C. (1986), *SISKEP — A Software for the Human Translator*, Computer Aided Translation Project, Université Sains Malaysia, Penang, 6 p.

[15] TONG L. C. (1987) « The Engineering of a Translator Workstation », *Computers and Translation 2*, 11 p.

[16] WinSoft (1987) *Manuel d'utilisation de WinTool™, Version 1.1* WinSoft.

10

Intégration des connaissances en génie linguistique : niveaux, dimensions, objets et contraintes*

Jean-Yves MORIN

Université de Montréal, Montréal, Canada

En ce qui concerne leur langage,
les honnêtes gens (君子 *jūnzǐ*)
ne laissent rien au hasard.
Confucius, *Entretiens familiers*

Introduction

Problème fondamental du génie linguistique

Le problème fondamental du génie linguistique est l'élaboration d'environnements de travail pour les linguistes, les traducteurs et les rédacteurs permettant la manipulation intelligente d'objets sémiotiques de type linguistique[1].

* La recherche sous-jacente à ce travail a été rendue possible grâce à des subventions du Conseil de recherches en sciences naturelles et en génie du Canada, de l'Office de la langue française du Québec et de la Coopération France-Québec (ministère des Affaires internationales, Québec et ministère des Affaires étrangères, France) ainsi que grâce à des contrats de recherche avec le Centre canadien de recherches en informatisation du travail (CCRIT, ministère des Communications, Canada). Les opinions exprimées ici n'engagent évidemment que l'auteur.

1. Les objets sémiotiques de type graphique ou musical ne relèvent donc pas du génie linguistique, pas plus que les textes structurés selon des systèmes textuels articulés qui ne sont pas des langues naturelles (programmes, expressions mathématiques, formules chimiques, etc.).

De tels environnements devront disposer d'une quantité importante de connaissances linguistiques. Ces connaissances devront éventuellement être combinées à des connaissances extra-linguistiques, de nature encyclopédique ou pédagogique, par exemple[2].

Le volume et la disparité de ces connaissances linguistiques pose de délicats problèmes d'INTÉGRATION[3]. D'une part, ces connaissances sont traditionnellement partitionnées en NIVEAUX plus ou moins autonomes qu'il est essentiel de pouvoir interfacer les uns aux autres de façon efficace[4]. D'autre part, même à l'intérieur d'un niveau les REPRÉSENTATIONS utilisées pour caractériser les connaissances linguistiques peuvent être très différentes.

Ainsi, pour caractériser l'*ordre des constituants* à l'intérieur d'un syntagme, une théorie comme la théorie gouvernement-liage (Chomsky (1981), dorénavant GL) utilisera essentiellement trois paramètres binaires (tête initiale/finale, assignation de Cas vers la gauche/droite et assignation de θ-rôle vers la gauche/droite)[5], tandis qu'une

---

Nous ne nous intéresserons ici qu'aux problèmes liés à la manipulation intelligente de « textes » au sens classique, c'est-à-dire d'objets structurés autour d'une représentation linéaire. C'est dire que nous ne tiendrons pas compte des développements en ce qui concerne les HYPERTEXTES structurés autour de représentations hiérarchiques ou hétérarchiques. Il est évident que les contraintes apportées par des systèmes dotés d'HYPER-MODÈLES ou d'HYPER-CRITIQUES changeront certainement de façon radicale la nature des activités de rédaction et de traduction dans les domaines où elles seront applicables. Voir à ce sujet Barrett, E. (1988) et (1989 réd.), par exemple.

Nous n'avons pas non plus tenu compte de la dimension stylistique et esthétique.

2. À titre d'exemple, mentionnons que les « connaissances linguistiques » intégrées aux traitements de texte universels actuels sont limitées à la détermination de segments linéaires : caractères, mots, phrases, paragraphes, pages, sections, etc. Un segment donné ne diffère d'un autre que par sa longueur, sa position et la séquence des marques qui le composent.

   Notons cependant qu'un bon traitement de texte dispose également de connaissances typographiques sur les en-têtes, les pieds de page, les marges, les alinéas, les espacements, l'orthographe (non grammaticale), la césure, la justification, les renvois et les notes infra-paginales, les titres, la hiérarchie des parties d'un texte, les types de caractères (polices, tailles, styles, corps), etc.

   Il faut distinguer les systèmes de traitement de texte universels éditeurs de structures (tels EMACS), qui, eux, sont dotés de certaines connaissances supplémentaires sur les types de structures qu'ils manipulent (programmes LISP dans le cas d'EMACS). La plupart des traitements de texte seront bientôt dotés de certaines connaissances structurales sur des types de textes standards (rapports, lettres, bulletins d'information, articles, bibliographies, etc.) et sur leurs divisions habituelles. Mais, ici encore, on touche à l'univers de l'hypertexte, plutôt qu'à celui de la langue naturelle.

3. Sans compter les problèmes D'ÉLABORATION. En effet, les connaissances explicites sur la nature et le fonctionnement du langage sont encore assez rudimentaires. Malgré certaines déclarations fracassantes, les progrès réalisés au niveau de la description des langues particulières et de la compréhension des mécanismes généraux du langage au cours des trente dernières années sont loin d'avoir été foudroyants.

4. Ce problème ne se pose généralement pas avec une telle acuité au linguiste, puisque ces différents niveaux correspondent plus ou moins aux sous-disciplines de la linguistique. Les descriptions se situent le plus souvent à l'intérieur d'un seul niveau. Voir cependant Zwicky (1986) pour un ensemble de travaux récents sur ce problème.

5. *Cf.* Travis (1989) et Dean Fodor & Crain (1990). Les paramètres en question ont des effets ailleurs que dans le domaine des contraintes d'ordonnancement. Ceci constitue une qualité pour les tenants de la théorie GL, puisque cela ajoute à la « richesse de la structure déductive ». Du point de vue de la représentation des connaissances linguistiques, c'est plutôt là un défaut, puisque cette propriété empêche de maintenir la transparence fonctionnelle des descriptions et donc la modularité des représentations.

théorie comme la grammaire syntagmatique généralisée (Gazdar *et al.* (1985), dorénavant GSG) utilisera un type particulier de règles, les PL-règles.

Par exemple, en français, les compléments apparaissent après la tête lexicale (verbe, nom, adjectif ou préposition) dont ils dépendent, alors qu'en japonais, ils apparaissent avant.

(1) (a) *Pinocchio frappe le chemin avec ses pieds de bois*[6]

(b) [P [SN *Pinocchio* SN]
[SV [V *frappe* V]
[SN *le chemin* SN]
[SP [Prép *avec* Prép]
[SN [Dét *ses* Dét]
[N *pieds* N]
[SX *de bois* SX]
SN]
SP]
SV]
P]

6. Pour des raisons de simplicité, nous avons choisi une représentation par crochets étiquetés indentée de la structure syntagmatique. Cette représentation utilise des étiquettes qui sont des abréviations des catégories utilisées en GSG plutôt que de celles utilisées en GL. Structuralement, elle correspond essentiellement à une représentation GSG.

Une représentation de la structure syntagmatique en GL serait nettement plus complexe que celle ci-dessus, mettant en jeu plusieurs niveaux intermédiaires (SC, SI, SD) et plusieurs catégories vides (Infl, Comp, Dét, etc.), mais, pour ce qui nous concerne ici (la distribution des catégories lexicales effectivement réalisées relativement aux catégories syntagmatiques), elle serait essentiellement congruente à la représentation GSG.

La représentation de la structure syntagmatique de (1) en GL serait, en simplifiant quelque peu, quelque chose comme (i).

(i) [SC [Comp Comp]
[SI [SD [Dét Dét]
[SN *Pinocchio* SN]
SD]
[Infl Infl]
[SV [V *frappe* V]
[SN *le chemin* SN]
[SP [Prép *avec* Prép]
[SD [Dét *ses* Dét]
[SN [N *pieds* N]
[SX *de bois* SX]
SN]
SD]
SP]
SV]
SI]
SC]

(2) (a) *Pinokkio-wa ki no ashi de michi-o keri.*
'*Pinocchio*-WA *bois*-NO *pied*-DE *chemin*-O *frapper*'

(b) [P [SN *Pinokkio-wa* SN]
[SV [SP
[SN
[SP *ki no* SP]
[N *ashi* N]
SN]
[Postp *de* Postp]
SP]
[SN *michi-o*]
[V *keri* V]
SV]
P]

En GL, on posera donc que le français est une langue à tête initiale, avec assignation de Cas et de θ-rôle vers la droite, alors que le japonais est à tête finale, avec assignation de Cas et de θ-rôle vers la gauche[7].

(3) Paramètres GL

(a) Français

Tête	Cas	Θ-rôle
Initiale	**Droite**	**Droite**

(b) Japonais

Tête	Cas	Θ-rôle
Finale	**Gauche**	**Gauche**

7. En fait, la description des contraintes d'ordre des mots en français (et en japonais) dans le cadre GL est beaucoup plus complexe que ne le laissent soupçonner ces tableaux. La valeur des trois paramètres est également paramétrisable. En français, par exemple, l'assignation du Cas accusatif par V ou oblique par Prép doit se faire vers la droite, mais l'assignation du Cas nominatif par INFL, vers la gauche, certaines têtes doivent être initiales, d'autres finales. On note par ailleurs que ce système contraint à une multiplication de niveaux, puisqu'il exclut *a priori* des structures « plates » où la tête est enrobée de part et d'autre de constituants (spécificateurs d'un côté et compléments de l'autre, par exemple [SN Dét N SP] ou [SV Aux V SN] en français), sauf dans une langue où le paramètre de tête (initiale/finale) ne serait pas actif.

Dans le cadre GSG, on posera que ce sont les CATÉGORIES LEXICALES (têtes, mais également spécificateurs — déterminants, auxiliaires, modificateurs, etc.) qui doivent apparaître avant les catégories syntagmatiques en français et après ces dernières en japonais.

(4) (a) Français

PL : [Niveau : **lexical**] ≪ ¬[Niveau : **lexical**]

(b) Japonais

PL : ¬[Niveau : **lexical**] ≪ [Niveau : **lexical**]

Si l'on tentait de calquer la théorie GL en GSG, on pourrait postuler un trait [TÊTE : α] caractérisant les têtes de syntagmes[8] et avoir des PL-règles comme en (5).

(5) (a) Français

PL : [Tête : α] ≪ ¬[Tête : α]

(b) Japonais

PL : ¬[Tête : α] ≪ [Tête : α]

On voit tout de suite la difficulté à intégrer une description du SN (français ou du japonais) suivant le cadre GSG à une description de la proposition conforme à GL, par exemple.

On pourrait penser que le problème de l'intégration des REPRÉSENTATIONS peut être évité en choisissant une fois pour toutes une théorie et en s'y tenant[9].

8. En fait, Zwicky (1988) postule quelque chose de semblable. Dans le système de Gazdar *et al.* (1985), ce trait serait, pour les têtes lexicales, une abréviation de [SOUS-CAT : n], où *n* est l'index d'une règle. Quant aux têtes non lexicales, si elles peuvent porter le trait [SOUS-CAT : n], comme le proposent Pollard et Sag (1987), elles sont identifiables de façon analogue. Sinon, il faudrait construire un trait *ad hoc*. Cependant, on remarque que leur comportement est différent de celui des têtes lexicales. Par exemple, les têtes lexicales précèdent leurs compléments, mais SV suit son sujet en français. C'est d'ailleurs ce qui nous amène à exprimer les contraintes d'ordre selon le NIVEAU (lexical ou non lexical) plutôt que la fonction (tête *vs* spécificateurs ou compléments).
9. Quant au problème de l'intégration des NIVEAUX, il demeure, de toute façon, incontournable.

L'expérience montre que tel n'est pas le cas[10]. La description de fragments importants de plusieurs langues naturelles présuppose la possibilité de réinterpréter les représentations de diverses théories dans la théorie choisie. De plus, même à l'intérieur d'une théorie donnée, la dynamique de la description entraîne des changements non négligeables de la théorie sous-jacente. Pourquoi dès lors ne pas faciliter directement cette réinterprétation?

Représentation des connaissances linguistiques

Objets et contraintes

Il existe deux grands types de représentation des connaissances, qu'il s'agisse de connaissances linguistiques ou non : les REPRÉSENTATIONS PAR OBJETS et les REPRÉSENTATIONS PAR CONTRAINTES.

Les REPRÉSENTATIONS PAR OBJETS définissent des univers d'objets, dotés de propriétés, entrant dans des relations et, éventuellement, capables d'actions (acteurs). Ces représentations mettent l'accent sur l'autonomie des objets, sur leur regroupement en classes organisées hiérarchiquement. Les objets d'une hiérarchie de classes partagent les uns avec les autres des propriétés héritées de leurs classes ancestrales communes. Ils peuvent prendre des valeurs par défaut ou déléguer à leurs ancêtres différentes fonctions. Cette conception semble intuitivement correspondre plus directement à notre ontologie naïve.

Les REPRÉSENTATIONS PAR CONTRAINTES définissent des univers de relations abstraites ou contraintes. Les individus ou objets n'y existent et n'y sont identifiables qu'à travers les contraintes qu'ils satisfont. Ce type de représentation favorise l'abstraction et l'inférence. Cette conception semble correspondre à une ontologie plus formelle.

Au niveau linguistique, on peut dire qu'un DICTIONNAIRE constitue une REPRÉSENTATION PAR OBJETS du langage, alors qu'une GRAMMAIRE en constitue plutôt une REPRÉSENTATION PAR CONTRAINTES.

Objets

Représentation à base d'objets : grammaires à base lexicale

Plusieurs théories grammaticales récentes[11] partent de l'hypothèse que les connaissances grammaticales (du moins celles qui sont spécifiques à une langue) sont essentiellement rassemblées dans le lexique. Cependant, outre un certain nombre de

10. Jusqu'à tout récemment, le groupe EUROTRA semblait favoriser un tel type d'approche. *Cf.* Arnold & des Tombe (1987).

11. Par exemple GL, où lexique et paramètres sont fondamentaux, la grammaire lexicale-fonctionnelle (dorénavant GLF) ou la grammaire syntagmatique endocentrique (*head-driven phrase structure grammar*, dorénavant GSE), Pollard (1984), Sag & Pollard (1987).

remarques et de travaux intéressants, mais de nature plutôt programmatique, sur le rôle et le fonctionnement du lexique, sur sa position relative face aux autres composantes, sur sa structure générale ou sur la forme des entrées lexicales, très peu de chercheurs se sont penchés sur la construction effective de lexiques explicites[12].

Dans les théories linguistiques basées sur la notion d'*information grammaticale*[13], le rôle du lexique est particulièrement important, puisque celui-ci y constitue le répertoire principal (tant par l'envergure que par la complexité) de l'information grammaticale spécifique à une langue.

De plus, dès que l'on se préoccupe de la *couverture* d'une description, tout autant que de sa rigueur, les problèmes associés à la construction de vastes lexiques deviennent incontournables[14].

En traitement des langues naturelles, d'autres facteurs encore contribuent à donner au lexique un rôle primordial. Le lexique contient l'ensemble des objets linguistiques directement accessibles et donc de coût constant, en ce qui concerne les ressources. Les autres objets (structures, relations) ne sont accessibles que par construction[15] et sont donc de coût variable (éventuellement prohibitif) toujours en ce qui a trait aux ressources. En gros, il est plus facile de vérifier une information que de la (re)construire. Ce fait est reconnu, au moins implicitement, par la plupart des chercheurs du domaine.

À un niveau abstrait, un lexique constitue une *représentation par objets* des connaissances grammaticales. C'est un ensemble d'objets autonomes appartenant à des classes organisées en des réseaux plus ou moins serrés (par exemple, catégories, catégories lexicales, catégories lexicales majeures ou verbes, verbes transitifs, verbes transitifs à double objet, etc.). Les objets des sous-classes héritent des propriétés de leurs superclasses (ou leur délèguent des propriétés). L'information sur le comportement syntaxique d'un objet (par exemple, la transitivité d'un verbe) est stockée dans la classe de l'objet (par exemple, du verbe) et n'a pas à être répétée dans l'entrée lexicale de celui-ci. Si un verbe appartient à la classe des verbes transitifs, il aura un comportement syntaxique de verbe transitif, dont les détails sont fixés une fois pour toutes et n'ont pas à être répétés.

Ce type de représentation permet de construire des descriptions de façon très modulaire, incrémentale et dynamique et d'exprimer des valeurs par défaut.

On peut aller jusqu'à poser l'hypothèse que toute la connaissance grammaticale peut être représentée en objets.

Non seulement les objets lexicaux sont associés à des règles (ou des familles de règles[16]), mais les règles elles-mêmes peuvent être vues comme des objets abstraits.

---

12. Notons cependant les travaux de Gross (1975) sur les *lexiques-grammaires* et ceux de Mel'čuk (1984, 1988) sur les *dictionnaires explicatifs et combinatoires* (DEC) qui, malgré leur intérêt, s'intègrent difficilement aux théories syntaxiques issues de la grammaire générative.
13. Sur cette notion, *cf.* Sag & Pollard (1987).
14. Sur le problème de la couverture, *cf.* Bouchard, Emirkanian et Morin (1991).
15. Si l'on compare une analyse à une démonstration, les entrées lexicales constituent autant d'axiomes, directement démontrés, alors que les structures ne peuvent être démontrées que de façon constructive, à partir des règles du système, ce qui constitue une source de complexité calculatoire.
16. *Cf.* Morin (1989), Blache (1990), Blache & Morin (1990) sur différentes notions de *famille de règles*.

Les règles et les principes sont organisés en réseaux de classes et de sous-classes (correspondant en partie aux « modules » d'une théorie comme GL ou GSG). Elles ont des propriétés qui peuvent être héritées (ou déléguées).

Ainsi, une règle ou un principe en GL a différents attributs : une strate ou un ensemble de strates d'application (DS, SS, FL, etc.), des paramètres, etc.

Par exemple, l'assignation de Cas en GL peut être vue comme une classe d'objets ayant des paramètres de CAS (eux-mêmes formés d'une VALEUR et d'un ASSIGNATEUR), d'une DIRECTION et d'une STRATE d'application (qui serait, par défaut la *S-tructure*).

(6) Assignation de Cas (GL)

$$\left[\begin{array}{ll} \text{CAS :} & \left[\begin{array}{ll} \text{VALEUR :} & c \\ \text{ASSIGNATEUR :} & \left\{\begin{array}{c} V \\ Infl \\ Prép \\ \ldots \end{array}\right\} \end{array}\right] \\ \text{DIRECTION :} & \left\{\begin{array}{c} \textit{gauche} \\ \textit{droite} \end{array}\right\} \\ \text{STRATE :} & s \textit{ (S-structure)} \end{array}\right]$$

De même, en GSG, une DI-règle a une IDENTITÉ formée d'un NOM ET d'un INDEX ET d'une FAMILLE, une FONCTION — décomposition, adjonction, coordination —, et un CONTENU, CONSTITUÉ d'une MÈRE (ou PARTIE GAUCHE), d'un ensemble de FILLES (ou PARTIE DROITE), qui peuvent être des FILLES-TÊTES (dont le NIVEAU est, par défaut, lexical), des FILLES-COMPLÉMENTS, des FILLES-SPÉCIFICATEURS ou des FILLES-ADJOINTES. La présence de tous ces attributs est héritée, pour chaque DI-règle, du fait même qu'il s'agisse d'une DI-règle.

(7) DI-règle (GSG)

$$\left[\begin{array}{ll} \text{IDENTITÉ :} & \left[\begin{array}{ll} \text{NOM :} & n \\ \text{INDEX :} & i \\ \text{FAMILLE :} & f \end{array}\right] \\ \text{FONCTION :} & \left\{\begin{array}{c} \textbf{décomposition} \\ \textbf{adjonction} \\ \textbf{coordination} \end{array}\right\} \\ \text{CONTENU :} & \left[\begin{array}{ll} \text{MÈRE :} & p \\ \text{FILLES :} & \left[\begin{array}{l} \text{FILLES-TÊTES : } t \text{ ([NIVEAU : } \textbf{lexical}\text{])} \\ \text{FILLES-COMPLÉMENTS : } c \\ \text{FILLES-SPÉCIFICATEURS : } s \\ \text{FILLES-ADJOINTES : } a \end{array}\right] \end{array}\right] \end{array}\right]$$

Les GRAMMAIRES D'OBJETS constituent donc un métaformalisme où différents types de connaissances grammaticales (correspondant, par exemple, à différentes théories linguistiques) peuvent être exprimés.

L'élaboration de (méta)formalismes de *représentation par objets* de la connaissance grammaticale (les GRAMMAIRES D'OBJETS) constituera un des objectifs fondamentaux de la recherche en génie linguistique dans les années à venir.

Contraintes

Représentations à base de contraintes : grammaires de règles et grammaires de principes

On oppose souvent les théories grammaticales *basées sur des règles* (comme GSG ou GLF) et les théories *basées sur des principes* (comme GL)[17].

En fait, les grammaires de principes comme les grammaires de règles (ainsi que les grammaires traditionnelles d'ailleurs) constituent des *représentations par contraintes* des connaissances grammaticales.

L'avantage de telles représentations est qu'elles permettent l'abstraction sur les objets et leur représentation. On peut exprimer des contraintes sans avoir à faire référence aux objets spécifiques qui leur seront soumis.

Les contraintes s'appliquent à des variables qui n'ont pas à être instanciées, mais uniquement typées abstraitement[18].

Ainsi, en formulant une notion comme celle de *c-commande* en GL, on n'a pas à se soucier de la façon précise dont seront représentées les structures syntaxiques ni des étiquettes sur les nœuds.

(8)

C-COMMANDE

α *c-commande* β si et seulement si α et β ne se dominent pas l'un l'autre et que le premier nœud branchant qui domine α domine également β.

Dans cette définition, α et β sont implicitement typés comme des nœuds (étiquetés) dans une structure arborescente σ, également implicite.

17. En fait, cette opposition est plus méthodologique que conceptuelle. Méthodologie *ascendante* : du spécifique au général dans le cas des *grammaires de règles vs* méthodologie *descendante* : du général au spécifique, dans le cas des *grammaires de principes*. La notion de *principe* est malheureusement assez floue et sert parfois à masquer une absence de rigueur et une confusion entre spécifique (qui s'oppose à général) et explicite (qui s'oppose à implicite). La concision, la clarté ou la simplicité peuvent constituer des qualités d'une description, qui peut entraîner un peu d'implicite. Mais l'implicite n'est jamais une qualité en soi (du moins en science).
18. D'ailleurs, toute contrainte constitue elle-même une fonction de typage abstrait.

On a donc la définition suivante :

(9)

C-COMMANDE

c-commande (α, β, σ) **ssi** (*PNBD* (γ, α, σ) $\supset$ *domine* (γ, β, σ))

Cette définition présuppose celle d'un prédicat *domine*, dont les trois arguments (α, β et σ) sont des nœuds étiquetés (α et β) et une structure (σ) et qui lui-même présuppose un prédicat *sous-arbre*, qui peut être défini à partir de primitifs d'égalité, d'appartenance et de représentation d'arbres[19] :

(10)

DOMINE

domine (α, β, σ) **ssi**
sous-arbre ($\alpha(\lambda)$, σ)
et
sous-arbre ($\beta(\lambda')$, $\alpha(\lambda)$)
et
$\neg$*égal* ($\alpha(\lambda)$, $\beta(\lambda')$)

(11)

SOUS-ARBRE

sous-arbre (σ, $\alpha(\lambda)$) **ssi**
(*égal* (σ, $\alpha(\lambda)$)
ou
(*membre* (σ', λ)
et
sous-arbre (σ, σ')))

19. Dans cette représentation, le terme $\alpha(\lambda)$ est une abréviation pour un arbre étiqueté dont α est la racine étiquetée et dont λ est la liste (éventuellement vide) des sous-arbres immédiats.

Les variables à gauche du connecteur **ssi** sont quantifiées universellement et les variables non liées à sa droite sont quantifiées existentiellement (*i.e.* de façon constructive).

Les prédicats *branchant* et *premier nœud branchant dominant* (*PNBD*) sont également implicites.

(12)

BRANCHANT

branchant (α, σ) **ssi** *sous-arbre* ($\alpha(\lambda)$, σ)
et
longueur (λ, 1)[20]
et
$1 > 1$

(13)

PNBD

PNBD (α, β, σ) **ssi** *domine* (α, β, σ)
et
branchant (α, σ)
et
(*domine* (γ, α, σ)
ou
$\neg$ (*domine* (γ, β, σ)
et
branchant (γ, σ)))

On peut éliminer cette définition un peu lourde en tenant compte de la transitivité de la relation de dominance, qui entraîne que si le premier nœud branchant qui domine α domine également β, alors *tous* les nœuds branchants dominant α dominent β.

Dès lors on n'a plus besoin, pour définir *c-commande*, du prédicat *PNBD*, mais uniquement des prédicats *domine* et *branchant*.

(14)

C-COMMANDE

c-commande (α, β, σ) **ssi** ((*branchant* (γ, σ)
et
domine (γ, α, σ))
$\supset$
domine (γ, β, σ))

20. Le prédicat *longueur* relie un premier argument de type *séquence* à un deuxième argument de type *nombre entier* représentant la longueur de cette séquence.

C'est donc sur le domaine (infini) des arbres étiquetés que porte la contrainte.Cette définition est purement relationnelle et totalement indépendante de la nature et de la représentation précise des objets qu'elle concerne.

On retrouve les mêmes possibilités d'abstraction sur les données dans des approches ascendantes, basées sur des règles générales. Par exemple, en GSG, une PL-règle comme (7) est tout à fait indépendante des représentations choisies, tant pour les structures catégorielles que pour les arbres.

(15)

[Spéc : α] $\ll$ α[21]

Contraintes et domaines

Le problème fondamental de l'intégration des descriptions linguistiques est qu'il n'existe pas de métalangage commun aux différentes théories linguistiques actuelles. Il est donc difficile de pouvoir évaluer exactement quelles sont les contraintes qu'elles apportent sur ce que Milner (1989) appelle le « possible de langue », c'est-à-dire comment elles limitent effectivement la notion de « langue naturelle possible ».

Trop souvent, on confond la notion de contraintes sur les « langues naturelles possibles » et celle, beaucoup moins intéressante, de contraintes sur les « descriptions possibles » de langues naturelles. Si l'on contraint les descriptions des langues naturelles à être exprimées dans un système n'utilisant que l'alphabet romain, par exemple, excluant ainsi les caractères grecs, ou à n'être rédigées qu'en latin classique, on apporte

---

21. *Cf.* Morin (1989).

Ici α subsume MAJEUR : β / NIVEAU : γ / FORME : δ et β, γ et δ sont dans la portée de MAJEUR, NIVEAU et FORME respectivement (*i.e.* $\beta \in \rho$(MAJEUR), $\gamma \in \rho$(NIVEAU) et $\beta \in \rho$(FORME)) la portée (ρ) est définie pour tous les traits et non seulement pour les traits à valeur atomique.

Le prédicat théorique PL, correspondant à l'opérateur $\ll$ peut être défini de la façon suivante :

PL (α, β, $\gamma(\lambda)$) **ssi** $\neg$*précède* (β, α, λ)

Le prédicat *précède* peut à son tour être défini récursivement en termes des primitifs *égal* et *membre*.

précède (α, β, λ) **ssi** ((*égal* (λ, $\alpha(\xi).\lambda'$) **et** *membre* ($\beta(\chi)$, λ')) **ou** (*égal* (λ, $\gamma(\varphi).\lambda'$) **et** $\neg$*égal* (β, γ) **et** *précède* (α, β, λ')))

Le point dans $\alpha(\xi).\lambda'$ et $\gamma(\varphi).\lambda'$ est l'opérateur de concaténation de séquence (le CONS de liste). Si la séquence $\lambda = \alpha.\lambda'$, α est le premier élément de la séquence λ et λ' le « reste » de cette séquence.

des contraintes, très fortes, sur la forme des descriptions possibles, mais absolument vides quant au contenu desdites descriptions.

Les contraintes sur la forme et le fonctionnement des grammaires n'ont de sens que si elles ont des conséquences directement déductibles sur la classe des objets (langages) décrits, sinon il ne s'agit que de variantes notationnelles. Ainsi, pour prendre un autre exemple, un peu plus technique, le fait de poser des contraintes très fortes sur les règles d'une grammaire indépendante du contexte telles que :

(16) (a) les règles sont de la forme (forme normale de Chomsky, FNC)

$A \longrightarrow B\ C$ ou $A \longrightarrow a$

ou encore

(b) les règles sont de la forme (forme normale de Greibach, FNG)

$A \longrightarrow a\ \varphi$ ou $A \longrightarrow a$

où A, B et C $\in V_n$ (le vocabulaire non terminal de la grammaire),
a $\in V_t$ (le vocabulaire terminal de la grammaire) et
$\varphi \in (V_n \cup V_t)^*$

ne change en rien la classe des objets que l'on peut décrire, même si cela réduit radicalement la classe des grammaires possibles. Pour ce qui est des objets qui peuvent être décrits, il s'agit toujours de la classe tout entière des langages de type 2. Si la réduction d'une grammaire à l'une ou l'autre de ces formes normales peut avoir un intérêt (pour l'écriture d'algorithmes de reconnaissance d'un certain type, ascendant pour FNC, prédictif pour FNG, par exemple), ce n'est certainement pas un intérêt théorique.

Comment comparer, par exemple, les hypothèses sur le liage[22] d'une théorie comme gouvernement-liage, et celles d'une théorie comme la grammaire syntagmatique généralisée[23] ? Il faudrait pouvoir définir les théories en fonction des *contraintes* sur certains *domaines* définissables indépendamment de la théorie. C'est ce qu'ont fait Miller (1990) pour les *systèmes de liage*, Gazdar *et al.* (1988) pour les *théories catégorielles*, Pullum et Barker (1990) pour les relations de *commande* et nous-même pour différents domaines syntaxiques (Morin (1989)).

En traitement des langues naturelles, cette nécessité théorique s'associe à des exigences pratiques. Idéalement, on voudrait pouvoir pratiquer un certain éclectisme

22. Ou plus précisément, sur le *liage-filtrage*. Sur cette notion de *liage-filtrage* (ensemble des liens globaux entre objets présents et objets « absents » —filtés— dans la structure) voir Morin (1989).
23. En GL, les phénomènes sont décrits au moyen de l'interaction d'un réseau de notions extrêmement complexe et mouvant d'un auteur à l'autre : *chaînes, gouvernement, K-markage, θ-marquage, filtres, principe de projection, θ-critère*, A et ¬A-*positions*, etc. En GSG, on fait appel à des principes universels de propagation de valeurs (essentiellement le *principe de traits de pied* (Gazdar *et al.* (1985)) ou le *principe de traits de liage* (Sag & Pollard (1987))) associés à un système de traits typés et à des restrictions intracatégorielles (restrictions de cooccurrence de traits, spécifications de traits par défaut) et extracatégorielles (DI-règles) sur les valeurs de certains traits.

éclairé, qui permettrait d'importer dans un système $\mathbb{S}$ les contraintes exprimées par la théorie $\mathbb{A}$ sur un domaine $\mathbb{D}$, de les combiner avec les contraintes exprimées par la théorie $\mathbb{B}$ sur un domaine $\mathbb{D}'$ (non nécessairement entièrement distinct de $\mathbb{D}$) et avec les contraintes exprimées par la théorie $\mathbb{C}$ sur le domaine $\mathbb{D}''$ et ainsi de suite.

Au niveau syntaxique, on pourrait, par exemple, disposer

(a) d'une représentation relativement adéquate des phénomènes $\mathbb{DV}$ de DIATHÈSE VERBALE en français (les « voix » de la grammaire traditionnelle : ACTIVE, PASSIVE, PRONOMINALE, NEUTRE, MOYENNE, CAUSATIVE) exprimée dans le cadre de la *grammaire lexicale-fonctionnelle* $\mathbb{GLF}$;

(b) d'une représentation des phénomènes de CLITICISATION PRONOMINALE $\mathbb{CP}$ dans le cadre de la théorie *gouvernement-liage* $\mathbb{GL}$;

et

(c) d'une représentation des phénomènes de COOCCURRENCE et d'ORDONNANCEMENT DES COMPLÉMENTS $\mathbb{COC}$ dans le cadre de la théorie de la *grammaire syntagmatique généralisée* $\mathbb{GSG}$.

Dès lors, on pourrait vouloir intégrer ces trois fragments dans un système $\mathbb{S}$:

(17)

$$\mathbb{S} = s(\mathbb{GLF}(\mathbb{DV})) \;\&\; s(\mathbb{GL}(\mathbb{CP})) \;\&\; s(\mathbb{GSG}(\mathbb{COC}))^{24}$$

Comment intégrer ces différentes descriptions, les rendre compatibles l'une avec l'autre et les exploiter dans un système efficace et transparent?

Ce que nous nous proposons de construire, c'est un cadre métathéorique général basé sur les notions de *dimensions*, d'*objets* et de *contraintes* permettant de réaliser à la fois l'œcuménisme théorique nécessaire et l'intégration des différents niveaux de représentation.

La définition de ces domaines et l'élaboration de (méta)formalismes d'expression de *contraintes* (les GRAMMAIRES DE CONTRAINTES) constitueront un autre objectif fondamental de recherche dans les années à venir.

24. s est ici une fonction générique d'interprétation de l'ensemble $\mathbb{T}(\mathbb{D})$ des énoncés de la théorie $\mathbb{T}$ sur le domaine $\mathbb{D}$ dans $\mathbb{S}$.

Problèmes d'intégration inter-niveaux : interfaces

Niveaux

C'est un fait bien connu que les connaissances linguistiques ne sont pas monolithiques, mais organisées en plusieurs articulations, modules ou niveaux. Chacun de ces niveaux porte sur un ensemble d'objets et de relations spécifiques, ce qui implique des représentations de types essentiellement différents. Or, même dans une conception en *relais* (comme celle de Jackendoff (1987)), où chaque niveau n'est interfacé qu'aux niveaux immédiatement voisins, plutôt qu'à l'ensemble des autres niveaux, l'intégration des niveaux pose autant de problèmes que la représentation du contenu de chacun d'entre eux.

En d'autres termes, le contenu des « flèches » dans un schéma comme le suivant est tout aussi complexe que celui des objets qu'elles relient.

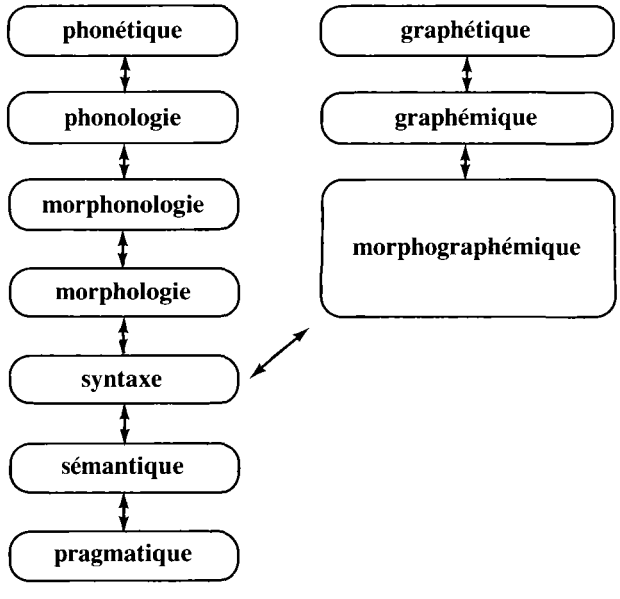


Figure 1 : Interaction des composantes dans une structure séquentielle

Par ailleurs, il semble bien que le lexique doive être interfacé à plusieurs niveaux différents, sinon à tous les niveaux de représentation linguistique.

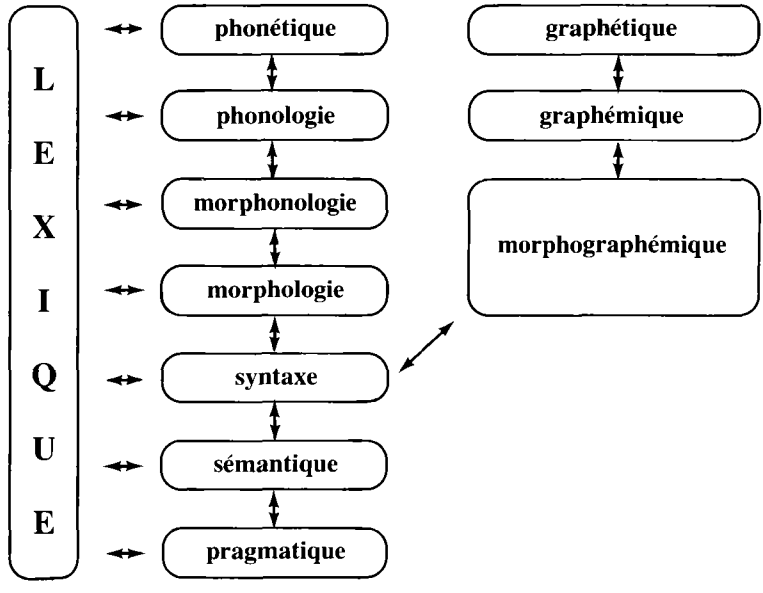


Figure 2 : Intégration du lexique aux autres composantes dans une structure séquentielle

Pis encore, dans un système hétérarchique, où chacun des niveaux peut interagir avec chacun des autres, comme les systèmes à dazibao, les problèmes d'interfaçage sont décuplés.

Comme le flot d'informations est beaucoup plus important entre modules voisins, on adopte généralement un modèle séquentiel à relais, plutôt qu'un modèle à dazibao, quitte à exprimer les autres contraintes au moyen d'extensions *ad hoc*.

Quoi qu'il en soit, l'immense majorité des systèmes expérimentaux existants ne porte que sur un ou deux de ces niveaux. Il y a donc là un gigantesque travail préliminaire à accomplir.

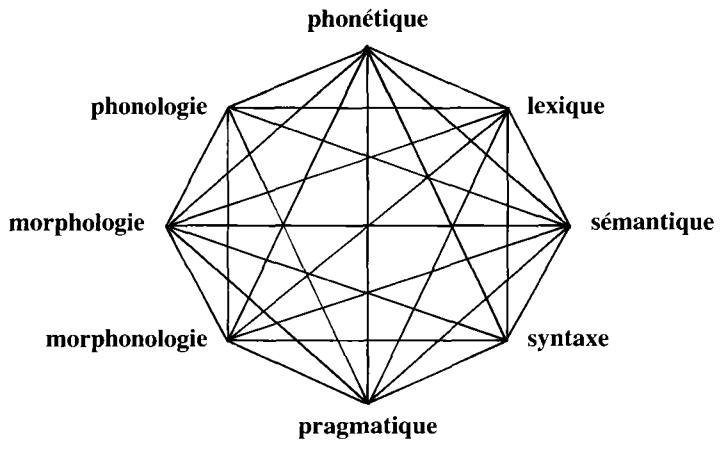


Figure 3 : Intégration des composantes dans une structure hétérarchique

Objets et interfaces

C'est ici que la méthodologie des représentations-objets est particulièrement intéressante. On peut concevoir chaque module comme une classe d'objets. Les classes sont elles-mêmes des objets qui communiquent avec les autres objets au moyen de messages. La façon dont un message est interprété par son récepteur dépend uniquement du contenu du message et des propriétés intrinsèques du récepteur.

On peut donc, en principe, définir à l'intérieur même d'un module à quels types de messages les objets du module peuvent réagir et comment ils peuvent y réagir.

Nous en sommes actuellement à concevoir des maquettes réalisant ce type d'interfaçage.

Intégration inter-représentations

Représentation des connaissances linguistiques et représentations de représentations

Métalangage de représentation

Toute théorie linguistique définit, implicitement ou explicitement, un métalangage de représentation des connaissances linguistiques doté d'une syntaxe et d'une sémantique propres.

Syntaxe

La syntaxe d'un métalangage de représentation des connaissances linguistiques est l'ensemble des expressions bien formées de ce métalangage.

Si, comme c'est le cas pour la théorie gouvernement-liage par exemple, le métalangage n'est pas explicitement défini, mais est construit à partir d'une langue naturelle quelconque, il est nécessaire de reconstruire ce métalangage de façon formelle.

Il s'agit tout d'abord d'établir le vocabulaire, c'est-à-dire les prédicats théoriques explicites de la théorie (par exemple *C-COMMANDE* , *GOUVERNE*, *ASSIGNE UN CAS*, *ASSIGNE UN θ-RÔLE*, etc.) et les prédicats théoriques implicites sous-jacents à la définition de ceux-ci (par exemple, *domine*, *est étiqueté*, *précède*, etc.) et de construire une syntaxe explicite et rigoureuse pour ceux-ci.

À titre indicatif, voici une typologie naïve et non exhaustive, des prédicats explicites communs à différentes variantes de la théorie gouvernement-liage et des modalités pouvant les affecter.

Dans une théorie comme la grammaire syntagmatique généralisée, qui est basée sur un ensemble de prédicats théoriques explicites (prédicats élémentaires : *SUBSUME*, *UNIFIE*, *a pour PORTÉE*, *a pour VALEUR*, *est une EXTENSION*, *une ABSTRACTION*, *une FUSION*, *une GÉNÉRALISATION*, prédicats-règles : *DI*, *PL*, *MÉTA*, *RCT*, *STD*, et prédicats principes universels : *PTT*, *PTP*, *PAC*), cet inventaire des prédicats théoriques sous-jacents est évidemment simplifié[25], comme est simplifiée la détermination de leur signature.

À ces prédicats, il faut ajouter les connecteurs logiques : *non* ($\neg$), *et* ($\wedge$), *ou* ($\vee$), *implique* ($\supset$) et *si et seulement si* (ssi, <=>), munis de leur interprétation habituelle. Dans cette liste de prédicats, la première forme entre parenthèses est une notation préfixée de la relation, la deuxième est la notation généralement employée[26].

---

25. *Cf.* Morin (1989) et (1989a) sur les prédicats théoriques de GSG et leur exportation dans des domaines d'interprétation externes.
26. Les symboles employés ont la signification suivante :

(i) α, α', β, β' (avec ou sans indice) sont des **catégories**, c'est-à-dire des ensembles de **spécifications de traits** $\{[t_i : v_i], [t_j : v_j], [t_k : v_k] \ldots\}$;

(ii) $\mathbf{t}$ est un **trait**, $\mathbf{t} \in \mathbf{T}$ (un ensemble fini de symboles, les **traits**);

(iii) $\mathbf{v}$ est une **valeur**, $\mathbf{v} \in \mathbf{V}$ (un ensemble fini de **valeurs**, atomiques ou complexes),
$\mathbf{V} = \mathbf{V_{ATOMIQUES}} \cup \mathbf{V_{COMPLEXES}}$
$\mathbf{V_{ATOMIQUES}}$ = un ensemble fini de symboles, disjoint de $\mathbf{T}$, les **valeurs atomiques**
$\mathbf{V_{COMPLEXES}}$ = l'ensemble $\mathbf{C}$ des catégories
$\mathbf{V'}$ est un sous-ensemble de $\mathbf{V}$ ($\mathbf{V'} \in \wp(\mathbf{V})$);

(iv) $\mathbf{W}$ est une variable pour un multi-ensemble quelconque de catégories (c'est-à-dire un ensemble où un même élément peut apparaître plusieurs fois. Par exemple, A_m = {a, b, a, c}$_m$ est un multi-ensemble construit sur l'ensemble de base A = {a, b, c});

(v) $\wp(\mathbf{V})$ est l'ensemble des sous-ensembles de $\mathbf{V}$.

PRÉDICATS

C-Commande
c-commande, appartient au domaine de
c-commande de

ACCESSIBILITÉ
est accessible, adjacent à

GOUVERNEMENT
gouverne,
est une catégorie gouvernante/de gouvernement (pour)

CONTRÔLE
contrôle

LIAGE
A/~A–lie, est A/~A–lié, est A/~A–libre,
est en relation de liage mutuel avec,
est une catégorie de liage,

est { *coindicé* / *coréférentiel* / *cosuperindicé* } *à*

est { *une anaphore* / *un pronominal* / *une variable* / *une r-expression* }

CONSTITUANCE, DOMINATION, PROJECTION
domine, contient, est un constituant immédiat de,
est une projection

SÉLECTION
sélectionne

POSITIONS
est / occupe une A/~A, θ/~θ, K/~K–position,
est une position génitive / prénominale

NIVEAUX
est lexical, est un X^O, X^1

BORNAGE, BARRIÈRES
est une borne / barrière

CAS, Θ-RÔLES
assigne un Cas / θ-rôle à, K-/θ-marque,
reçoit un Cas/θ-rôle de,
est θ-marqué par, absorbe le Cas (C) (de),
retire la possibilité d'assigner le Cas à

CHAÎNES
est une chaîne / un chemin ambigu/non ambigu, brise une chaîne / un chemin

VISIBILITÉ, MORPHOLOGIE, PHONÉTIQUE
est visible / doté d'une matrice phonétique, possède une matrice de traits intègre,
est un AGR « riche », etc.

EFFACEMENT DANS COMP
efface ($[_{N^-}$ *e]*), *assigne un indice à (Comp).*

MODALITÉS

NIVEAUX
en D-/S-structure, en FL, en FP,
en structure de surface,
à tous les niveaux de dérivation, au niveau N_i

GOUVERNEMENT
par gouvernement

DOMAINE LINGUISTIQUE
dans la (famille de) langue(s) L_j

LOCALITÉ
localement

MODALITÉS DIVERSES
lexicalement
structuralement
thématiquement
facultativement
proprement
minimalement/maximalement
directement/indirectement

Figure 4 : Prédicats et modalités de la théorie GL

Prédicats élémentaires

a) *a pour portée*	(PORTÉE [t, V']	ou	ρ (t) = V'	),
b) *a pour valeur*	(VAL [t, v].	ou	[t : v]	),
c) *subsume*	(SUBS [α, α'],	ou	$\alpha \subseteq \alpha'$	),
d) *est une extension de*	(EXT [α, α'],	ou	$\alpha \supseteq \alpha'$	),
e) *unifie*	(UNI [α, {β_1, β_2, ..., β_n}]	ou	(α, {β_1, β_2, ..., β_n})	),
f) *généralise*	(GÉN [α, {β_1, β_2, ..., β_n}]	ou	(α, {β_1, β_2, ..., β_n})	),
g) *est une abstraction de*	(ABS [α, {β_1, β_2, ..., β_n}]	ou	Λ (α, {β_1, β_2, ..., β_n})	),
h) *est une fusion de*	(FUS [α, {β_1, β_2, ..., β_n}]	ou	Φ (α, {β_1, β_2, ..., β_n})	).

Figure 5 : Prédicats élémentaires de GSG

Prédicats-règles

a) *domine immédiatement*	(DI [α, {β_1, ..., β_n}]	ou	$\alpha \rightarrow \beta_1, \ldots, \beta_n$	),
b) *précède linéairement*	(PL [α, β]	ou	$\alpha \ll \beta$	),
c) *RCT*	(RCT [α, β]	ou	RCT : $\alpha \supset \beta$	),
d) *STD*	(STD [α, β]	ou	STD : $\alpha \supset \beta$	),
e) *métarègle*	(MÉTA $\left[\begin{array}{l} <\alpha, <\beta, W \gg, \\ <\alpha', <\beta', W \gg \end{array}\right]$	ou	$\begin{array}{ll} \alpha & \beta, W \\ \alpha' & \beta', W \end{array}$	).

Figure 6 : Prédicats-règles de GSG

Principes universels d'instanciation

a) *Principe de traits de tête*
(PTT [α, {β_1, ..., β_n}, {β_1, ..., β_k}] ou PTT [α, (β_1, ..., β_1, ..., β_k, ..., β_n)]).

b) *Principe de traits de pied* (PTP [α, {β_1, ..., β_n}] ou PTP [α (β_1, ..., β_n)]),

c) *Principe d'accord-contrôle* (PAC [α {β_1, ..., β_n}] ou PAC [α (β_1, ..., β_n)]).

Figure 7 : Prédicats-principes universels de GSG

Voici les signatures de ces différents prédicats (c'est-à-dire les types des arguments qu'ils prennent et, dans le cas des fonctions, les types des valeurs qu'elles retournent)[27].

Prédicats-élémentaires

PORTÉE :	$T \longrightarrow \wp(V)$	(fonction)
VALEUR :	$T \longrightarrow V$	(fonction)
SUBSUME :	$((T \longrightarrow V), (T \longrightarrow V))$	(relation)
EXTENSION :	$((T \longrightarrow V), (T \longrightarrow V))$	(relation)
UNIFIE :	$\wp(T \longrightarrow V) \longrightarrow (T \longrightarrow V)$	(fonction)
GÉNÉRALISATION :	$\wp(T \longrightarrow V) \longrightarrow (T \longrightarrow V)$	(fonction)
ABSTRACTION :	$\wp(T \longrightarrow V) \longrightarrow (T \longrightarrow V)$	(fonction)
FUSION :	$\wp(T \longrightarrow V) \longrightarrow (T \longrightarrow V)$	(fonction)

Prédicats-règles

DI :	$((T \longrightarrow V), \wp(T \longrightarrow V))$	(relation)
PL :	$((T \longrightarrow V), (T \longrightarrow V))$	(relation)
RCT :	$((T \longrightarrow V), (T \longrightarrow V))$	(relation)
STD :	$((T \longrightarrow V), (T \longrightarrow V))$	(relation)

Principes universels d'instanciation

$((T \longrightarrow V), \wp(T \longrightarrow V)) \longrightarrow ((T \longrightarrow V), \wp(T \longrightarrow V))$ (fonction)

Figure 8 : Signatures des prédicats théoriques de GSG

Sémantique

Mais un tel travail ne suffit pas, encore faut-il donner une sémantique descriptive rigoureuse à chacun de ces prédicats. Il faut définir les domaines d'objets caractérisés et les relations que les prédicats théoriques y définissent. En d'autres termes, il s'agit de typer les prédicats relativement aux objets des domaines qu'ils servent à définir.

Ce travail, loin d'être trivial, est absolument nécessaire, bien qu'il soit la plupart du temps négligé. En effet, les « modèles informatiques » des théories linguistiques ne partagent le plus souvent avec ces dernières qu'un sous-ensemble non interprété du vocabulaire théorique.

27. On aura noté que la PORTÉE, la VALEUR, l'UNIFICATION, la GÉNÉRALISATION, l'ABSTRACTION et la FUSION ne sont pas des relations, mais des opérations (fonctions), que nous transformons en relations en ajoutant comme argument leur résultat. En fait, chacun des prédicats théoriques peut être reconstruit comme une fonction caractéristique retournant l'ensemble des objets pour lesquels il est vrai dans une interprétation donnée. Pour plus de simplicité, nous n'avons pas donné cette reconstruction ici.

Un tel travail d'identification de la *signature* des théories linguistiques est actuellement en cours pour la théorie gouvernement-liage et pour la théorie syntagmatique généralisée.

Domaines et dimensions

L'intégration des représentations que nous proposons est une intégration sémantique plutôt que syntaxique. C'est-à-dire que c'est à partir des domaines définis par les différentes théories (métasémantique), plutôt qu'à partir de leur structure interne (métasyntaxe) que nous tentons de les intégrer. Pour ce faire, il est nécessaire de procéder à un repérage *a priori* des dimensions de l'espace définissable pour chacun des niveaux de façon à pouvoir vérifier comment chaque théorie particulière recouvre cet espace, comment elle y établit son domaine. On peut, par exemple, relativement indépendamment de toute théorie syntaxique ou sémantique, définir les dimensions des espaces syntaxique et sémantique pour les langues naturelles.

Dimensions syntaxiques

Le domaine syntaxique peut être défini par les dimensions suivantes.

Structure syntagmatique

CATÉGORIES

ORDONNANCEMENT

HIÉRARCHIE

Dépendances

GOUVERNEMENT	LIAGES
sélection	anaphore
rection	coréférence
	quantification
	accord
FILTRAGES	LIAGES-FILTRAGES
répulsion	contrôle
absorption	« mouvements »
haplologie	(*arguments, opérateurs, parasites*)
anaccord	
exclusions	

BORNAGES

Globaux/locaux Horizontaux/verticaux

Figure 9 : Dimensions syntaxiques

Dimensions sémantiques

Le domaine sémantique peut se définir par les dimensions suivantes.

θ-structure
- Rôles des arguments
  - AGENT
  - PATIENT
  - LOCATION
  - SOURCE
  - CIBLE
  - ...

ρ-structure
- Rôles discursifs
  - THÈME-RHÈME
  - PERTINENCE / CENTRALITÉ
  - CENTRE / PÉRIPHÉRIE
  - NOUVEAU / CONNU
  - EMPHATIQUE / RETIRÉ,
  - FOCUS / PRÉSUPPOSITION

μ-structure
- Modalités
  - NÉCESSITÉ / POSSIBILITÉ
  - OBLIGATION / PERMISSION
  - PROBABILITÉ / CERTITUDE

ι-structure
- Structure intensionnelle
  - QUANTIFICATION
  - DÉTERMINATION
  - MODULATION
  - LIMITATION

τ-structure
- Temps et aspect

ε-structure
- Espace et lieux
- Parties et limites

π-structure
- Structure locutoire et performative

Figure 10 : Dimensions sémantiques

Les limites d'espace nous empêchent d'expliciter ici cet inventaire de dimensions syntaxiques et sémantiques et de montrer comment chaque théorie syntaxique ou sémantique doit pouvoir y être interprétée.

Conclusion : institutions grammaticales

Au fond, ce que nous devrons essayer de caractériser, avant de pouvoir interfacer des REPRÉSENTATIONS, ce sont des INSTITUTIONS GRAMMATICALES, c'est-à-dire des espaces, des cadres généraux indépendants des représentations liées à des théories linguistiques particulières. Par ailleurs, ces INSTITUTIONS GRAMMATICALES[28], définies selon leurs dimensions, permettront également de mieux voir la

28. Sur la notion formelle d'*institution* en logique des programmes, voir Goguen & Burstall (1983, 1985, 1986).

nature de chacun des NIVEAUX de représentation et comment ils peuvent être interfacés à leur tour les uns aux autres.

Nous sommes encore bien loin de cet idéal, mais les développements récents sont particulièrement encourageants. Le génie linguistique (comme, avant lui, le génie logiciel) ne sortira du bricolage qu'au prix d'une réflexion approfondie et rigoureuse sur la sémantique des modèles qu'il utilise.

Références

ARNOLD D. et L. DES TOMBE (1987) : « Basic Theory and Methodology in EUROTRA », *in* Nirenburg, S. (1987 réd.) *Machine Translation, Theoretical and Methodological Issues*, Cambridge, Cambridge University Press, p. 114-135.

BARRETT E. (1988) : *Text, ConText and HyperText : Writing with and for the Computer*, Cambridge, MA, MIT Press.

BARRETT E. (1989 réd.) : *The Society of Text : Hypertext, Hypermedia, and the Social Construction of Information*, Cambridge, MA, MIT Press.

BLACHE Ph. (1990) : *L'analyse syntaxique dans le cadre des grammaires syntagmatiques généralisées : interprétations et stratégies*, thèse de doctorat inédite, faculté des sciences de Luminy, Université d'Aix-Marseille II.

BLACHE Ph. et J.-Y. Morin (1990) : « Bottom-Up Filtering : A Parsing Strategy for ID/LP Grammars », *in* Karlsson, F. (1990 réd.) *COLING-90, Proceedings*, Helsinki, Université d'Helsinki et New York, Association for Computational Linguistics.

BOUCHARD L., L. EMIRKANIAN et J.-Y. MORIN (1991) : « Détermination de la couverture dans les interfaces en langue naturelle », *in* Gauthier, G. (1991 réd.) *ICO-91*.

CHOMSKY N. (1981) : *Lectures on Governement and Binding*, Dordrecht, Foris.

DA SYLVA L. (1990) : *Un parseur inspiré de la théorie gouvernement-liage*, mémoire de M.A. inédit, département de linguistique, philologie et traduction, Université de Montréal.

DEAN FODOR J. & S. CRAIN (1990) : « Phrase Structure Parameters », *Linguistics and Philosophy*, 13, 6 : 619-659.

GAZDAR G., E. KLEIN, G. PULLUM et I. SAG (1985) *Generalized Phrase Structure Grammar*, Oxford, Blackwell et Cambridge, MA, Harvard Univ. Press.

GAZDAR G., G. PULLUM, R. CARPENTER, E. KLEIN, T. HUKARI & R. LEVINE (1988) : « Category Structures », *Computational Linguistics*, 14, 1 : 1-19. (Traduction française dans Torris T. et Ph. Miller (1990 réd.) *Formalismes syntaxiques pour le traitement automatique du langage naturel*, Paris, Hermès).

GOGUEN J.A. et R.M. BURSTALL (1983) : « Introducing Institutions », *in* Clarke, E. et D. Kozen (1983 réd.) *Logics of Programs*, Lecture Notes in Computer Science, 164, Berlin, Springer-Verlag, p. 221-256.

GOGUEN J.A. et R.M. BURSTALL (1985) : *Institutions : Abstract Model Theory for Computer Science*, Rapport CSLI-86-54, Stanford, Center for the Study of Language and Information, Stanford University.

GOGUEN J.A. et R.M. BURSTALL (1986) : *A Study in the Foundations of Programming Methodology : Specifications, Institutions, Charters and Parchments*, Rapport CSLI-85-30, Stanford, Center for the Study of Language and Information, Stanford University.

GROSS M. (1985) : *Méthodes en syntaxe*, Paris, Hermann.

JACKENDOFF R. S. (1987) : *Consciousness and the Computational Mind*, Cambridge, MA, MIT Press.

MEL'ČUK I. *et al.* (1984, 1988) : *Dictionnaire explicatif et combinatoire du français, Recherches lexico-sémantiques I et II*, Montréal, Presses de l'Université de Montréal.

MILLER Ph. (1990) : « Systèmes de liage », *in* Torris, T. et Ph. Miller (1990 réd.) *Formalismes syntaxiques pour le traitement automatique du langage naturel*, Paris, Hermès.

MILNER J.-C. (1989) : *Introduction à une science du langage*, Paris, Seuil.

MORIN J.-Y. (1989) : *Syntaxe*, département de linguistique et philologie, Université de Montréal.

MORIN J.-Y. (1989a) : « Prédicats théoriques et données externes. Syntaxe diachronique », *Revue canadienne de linguistique*, numéro spécial, *Linguistic Theory and External Evidence.*, 33, 4 : 443-75.

MORIN J.-Y. (1989b) : « Particules et parsage universel », in Weydt, H. (1989 réd.) *Sprechen mit Partikeln*, Berlin, De Gruyter, p. 713-728.

POLLARD C.J. (1984) : *Generalized Phrase Structure Grammars, Head Grammars and Natural Language*, thèse de Ph. D. inédite, Stanford University.

PULLUM G. et C. BARKER (1990) : « A Theory of Command Relations », *Natural Language and Linguistic Theory.*

SAG I. et C.J. POLLARD (1987) : *Information-Based Syntax and Semantics, Vol. I : Fundamentals*, Stanford, CSLI et Chicago University of Chicago Press.

SAINT-DIZIER P. (1989) : « Programming in Logic with Constraints for Natural Language Processing », *ACL Europe 4* : 87-94.

SAINT-DIZIER P. et P. SZPAKOWICZ (1989 réd.) : *Logic Programming and Logic Grammars*, Londres, Ellis Horwood.

TRAVIS L. (1989) : « Parameters of Phrase Structure », *in* Baltin, M. & A. Kroch (1989 réd.) *Alternative Conceptions of Phrase Structure*, Chicago, University of Chicago Press.

ZWICKY A. (1986) : *Interfaces*, Colombus : Ohio State University Working Papers in Linguistics, n° 32.

ZWICKY A. (1988) : « Direct Reference to Heads », *Folia Linguistica*, 22.3-4 : 397-404.

11

Poste de travail de traduction par l'informatisation d'un dictionnaire français-arabe

Saadia TALEB

Institut d'Études et de Recherches pour l'Arabisation (IERA), Université Mohamed V, Rabat, Maroc

La réalisation d'un poste de travail de traduction par l'informatisation d'un dictionnaire français-arabe compatible avec un traitement de texte arabisé grand public se situe dans le cadre d'un contrat entre l'ACCT et l'association Champollion. Ce contrat associe l'IERA en tant qu'intervenant apportant une compétence spécifique en matière de mise au point de bases de données lexicographiques et terminologiques multilingues.

Combinant différentes aides informatiques, le poste de travail pour le traducteur de français vers l'arabe a pour objectif d'automatiser certaines tâches connexes à l'acte traductionnel.

Ce poste est installé sur petit matériel, muni d'affichage graphique (Macintosh) et largement diffusé, et repose sur le principe de l'intégration d'un ensemble d'outils informatiques et de savoir-faire réellement existant :

- Un traitement de texte Wintext qui offre la possibilité d'associer plusieurs langues dans le même document et permet de synchroniser deux fenêtres dans lesquelles apparaîtront le texte source et sa traduction. Le traitement de texte dispose d'un correcteur orthographique français qui utilise un dictionnaire de 246 000 mots avec la possibilité de mettre au point son propre lexique ainsi que celle d'exploiter les données de documents externes crées par d'autres applications, notamment Winfile et WinTool.

- Un lemmatiseur (ODILE) pour trouver automatiquement la forme canonique d'un mot en évitant de forcer l'utilisateur à taper cette forme. Il permet de fabriquer, à partir d'une chaîne sélectionnée, un ensemble de clés de recherche, de les compléter si nécessaire « à la main » et de s'en servir pour accéder aux dictionnaires gérés par WinTool.
- L'accès à un dictionnaire informatisé.

La recherche des équivalents est une opération essentielle de l'acte traductionnel. L'informatisation de cette tâche augmente la productivité du traducteur. Cette productivité augmente d'autant plus que le dictionnaire est modifiable et peut être mis à jour constamment par l'utilisateur.

L'outil dictionnaire du poste de travail du traducteur est pris en charge par l'utilitaire WinTool qui est un accessoire de bureau disponible dans le menu « pamno » du Macintosh à partir de l'application Wintext. Cet utilitaire permet d'ouvrir deux bases simultanément.

L'outil dictionnaire mis au point dans le cadre de ce travail est un dictionnaire français-arabe de 15 000 termes spécialisés dans les domaines des finances : comptabilité et banque. Ces termes sont extraits de la base de données LEXAR.

Cette base de données lexicographiques et terminologiques réalisés à l'IERA sur un IBM 4341, constitue un dictionnaire multilingue géant et se présente comme un instrument de référence lexicographique et terminologique de première importance et par ses dimensions (quelque 700 000 fiches) et par la diversité des sources à partir desquelles a été puisée la matière linguistique. Cette matière à la fois complexe et hétérogène subit une transformation lors d'une opération d'analyse qui vise à représenter le contenu du document primaire sous une forme différente en respectant un certain nombre de règles prédéfinies. Ces règles représentent un nombre précis pour permettre la distribution de la matière linguistique selon des catégories prédéterminées aisément reconnaissables par la machine et assez générales pour que, quelle que soit la complexité du dictionnaire ou du lexique à stocker, ce dernier puisse s'intégrer à l'ensemble et être réduit aux mêmes catégories prévues.

La donnée LEXAR est de ce point de vue une entité indépendante ayant ces propres spécifications et pouvant être réutilisée pour différentes applications.

Pour réaliser cet outil dictionnaire sur l'utilitaire WinTool, il a fallu sélectionner les données LEXAR relatives aux domaines des spécialités choisies, les extraire de l'IBM central, les transférer sur micro-ordinateur PS/2, procéder à un décodage et à un réencodage selon le codage arabe utilisé sur Macintosh, et enfin, définir une base de données WinTool et y transférer les données.

Cette opération a montré la possibilité de tirer de la base de données LEXAR des dictionnaires spécialisés de 7 000 à 8 000 termes et leurs équivalents pouvant tenir sur une disquette de 800 Ko, ou de créer une collection de tels dictionnaires sur CD-ROM.

On peut donc personnaliser le poste de traduction en offrant des dictionnaires spécialisés selon les besoins de tel ou tel utilisateur.

Il s'agit là d'une première expérimentation et il va sans dire que les outils d'aides doivent être encore plus perfectionnés et que l'on peut intégrer d'autres aides infor-

matiques, notamment un vérificateur orthographique pour l'arabe, un système de conjugaison et un logiciel de fenêtrage performant capable de gérer efficacement et simultanément toutes ces applications. Ces logiciels doivent répondre à une double exigence : une convivialité extrême et une haute performance.

12

Traduction et terminologie : expérience et perspectives industrielles

Daniel BACHUT, René GERBER[1]

SITE, Maisons-Alfort; B'VITAL, Grenoble, France

Résumé

Le groupe SITE, leader en ingénierie documentaire (documentation, traduction, logistique, etc.), développe aujourd'hui plusieurs activités, qui se situent au cœur même des domaines de la lexicographie, de la terminologie et de la traduction. Dans cet article, nous décrivons ces différentes activités, puis nous présentons les conditions techniques nécessaires à la conception du poste de travail du rédacteur-traducteur de l'an 2001.

La traduction

Le département de traduction

Le département de traduction de SITE compte aujourd'hui 60 traducteurs, qui traitent plus de 110 000 pages par an. C'est un département à croissance rapide, conséquence des besoins de plus en plus pressants du marché sur ce créneau. Les langues les plus couramment traitées sont le français et l'anglais, et dans une moindre mesure l'allemand

1. Les auteurs tiennent à associer à la rédaction de cet article Mme B. Roudaud, M. J.-J. Perot et M. D. Maret, sans l'aide desquels il aurait été difficile de réaliser un tour d'horizon aussi complet des activités de la société.

et l'espagnol. Il s'agit essentiellement de traduction technique dans des domaines tels que l'aéronautique, l'informatique, la mécanique, etc.

Les traducteurs travaillent dans un environnement informatisé moderne. L'organisation informatique de base consiste en un réseau de PC/PS, connectés par Token Ring. Différents traitements de texte sont disponibles, le plus couramment employé étant le logiciel Word de Microsoft. Le poste de travail donne accès à la base terminologique multilingue PHENIX, présentée plus loin. La variété des besoins des clients amène également les traducteurs à travailler dans d'autres types d'environnement : Macintosh et, depuis peu, stations de travail sous UNIX.

L'objectif de qualité conduit les équipes à pratiquer souvent la lecture croisée en guise de révision. Plusieurs types de relecture peuvent être effectués : relecture linguistique (orthographique et stylistique), relecture technique (par un traducteur spécialisé ou un spécialiste du domaine). L'aspect terminologique est pris en compte depuis de nombreuses années, dans un souci de productivité, d'homogénéité, de qualité, et même de formation. À cet égard, on estime que, depuis l'utilisation de la base terminologique PHENIX, le temps de formation d'un traducteur à une langue de spécialité est passé de deux ans à six mois. Souvent le client impose l'utilisation d'une *terminologie approuvée*, définie par lui, que les traducteurs doivent s'engager à respecter.

Ajoutons que les demandes des clients peuvent être très variées. Certains ont des exigences telles, que la traduction consiste quasiment en la réécriture du document en langue cible (traduction de l'anglais simplifié en français « normal », par exemple). D'autres demandent, au contraire, une traduction très fidèle du point du vue stylistique, c'est-à-dire aussi proche que possible du texte source. Cela est justifié pour des documents appelés à être modifiés, et pour lesquels le client tient à conserver un certain parallélisme entre les textes source et traduit pour simplifier les mises à jour.

La traduction assistée par ordinateur

L'industrialisation du système ARIANE F → A

Début 1990, SITE a pris une part majoritaire dans le capital de la société grenobloise B'VITAL, spécialisée dans le développement de systèmes de traduction assistée par ordinateur (TAO), et dont l'une des réalisations majeures consiste dans le développement d'un système de traduction français-anglais de deuxième génération, appelé ARIANE F → A. Ce système utilise le générateur de systèmes de TAO ARIANE, mis au point par le laboratoire GETA (Groupe d'études pour la traduction automatique). Les linguiciels (grammaires et dictionnaires) d'ARIANE F → A ont été conçus pour traiter des textes techniques, en particulier dans le domaine aéronautique.

Un projet d'industrialisation d'ARIANE F → A est actuellement mené par les trois partenaires (SITE, B'VITAL et le GETA), avec le soutien du ministère de l'Industrie. Il a pour objectif de faire passer l'outil d'un état de prototype avancé à un

état industriel. Pour y parvenir, les partenaires ont préconisé différentes actions, que l'on peut regrouper suivant deux axes principaux :

- l'« habillage » du système, par la réalisation des outils connexes permettant de l'intégrer dans une chaîne documentaire industrielle;
- la mise en service du système dans des conditions réelles, en lui faisant traduire plusieurs milliers de pages. Évaluation du système en ce qui a trait à la qualité des traductions, au coût, au délai et au volume.

Habillage du système

ARIANE fonctionne sur *mainframe* IBM, et propose aux linguistes un environnement de développement très riche. En exploitation, l'utilisateur final voudra disposer de son traitement de texte habituel (sous MS-DOS par exemple) et bénéficier d'un grand confort au niveau de l'interface (menus déroulant, multifenêtrage, souris, etc.). Un premier moniteur d'exploitation a donc été réalisé sous MS-DOS, pour faciliter l'accès au système de TAO, transformé alors en serveur de traduction. Cet environnement permet à un utilisateur de lancer sa traduction de manière transparente, sans qu'il lui soit nécessaire de savoir où celle-ci a réellement lieu. De plus, la traduction étant déportée, l'utilisateur peut continuer à travailler sur son micro-ordinateur. Actuellement la traduction peut être lancée à partir de MS-DOS, ou directement sous Word par l'intermédiaire d'un jeu de touches de fonctions.

Un travail important a également été réalisé pour permettre la restitution du format du texte source dans le texte traduit, dans le but de diminuer sensiblement la charge de travail du post-éditeur. Une représentation « normalisée » des ordres de formatage et de structuration des textes a été définie et prise en compte par les processus linguistiques. Des pré et post-processeurs, complètement intégrés au nouveau moniteur, ont été réalisés pour passer de la forme des textes d'entrée à la forme « normalisée », et de cette forme à la forme de sortie. Une généralisation de cette représentation en SGML est prévue.

Évaluation du système et premiers résultats

Les premiers tests sur la documentation de l'avion FALCON 900 ont été réalisés sans la terminologie spécifique, pour des raisons de délai et de disponibilité. La présence dans les dictionnaires d'ARIANE F $\longrightarrow$ A de nombreux termes du domaine aéronautique (8000 environ) a cependant rendu possibles ces essais. Notons que la récupération automatique du glossaire spécifique au FALCON 900 (articles PHENIX) vient d'être achevée.

Les traducteurs ont considéré que les traductions obtenues étaient de bonne qualité. Les calculs réalisés sur les temps de post-édition donnent aujourd'hui une demi-heure par page révisée pour les cas les plus défavorables, permettant d'améliorer la productivité des traducteurs par un facteur deux (16 pages par jour et par traducteur au lieu de 8 pages classiquement). Il est certain que l'ajout de la terminologie permettra de réduire encore le temps de révision.

L'amélioration de la productivité (par diminution des délais) ne se traduit cependant pas par une diminution comparable des coûts, puisqu'à celui du réviseur, il faut ajouter le coût des ressources informatiques nécessaires, ainsi que le coût du système lui-même. Le coût de l'amortissement du système est faible, dans la mesure où un système de TAO de ce type ne peut intéresser qu'un client ayant des besoins importants (plusieurs dizaines de milliers de pages par an).

Ce sont les ressources informatiques (essentiellement du CPU) qui représentent aujourd'hui l'élément de coût le plus important pour une traduction réalisée avec ARIANE F ⟶ A. Cela tient au fait que l'implémentation est dépendante de l'architecture IBM 370. Remarquons que le coût de la seconde CPU peut varier du simple au quadruple suivant les centres de calcul (toutes choses étant égales par ailleurs). Ces premières conclusions (bonne qualité, productivité double mais coûts informatiques élevés) ont motivé SITE à préparer le projet EUROLANG.

Le projet EUROLANG

En collaboration avec plusieurs partenaires industriels et universitaires européens, SITE a déposé un projet EUREKA fin mars 1991. Le projet EUROLANG a pour objectif la réalisation d'un système de TAO multilingue fonctionnant sur cinq langues européennes (français, anglais, allemand, espagnol et italien). Pour construire ce système, on développera une boîte à outils pour le traitement automatique de la langue naturelle (TALN).

Le principe d'une boîte à outils donne la possibilité d'utiliser une stratégie *meccano* (*plug and play*), qui offre des garanties d'ouverture tant au niveau informatique (adéquation de chaque outil, possibilité d'assemblage) qu'au niveau linguistique (mise en jeu de plusieurs théories, stratégies définies par les utilisateurs de la boîte à outils).

Le projet est bâti autour d'un noyau de partenaires représentant les différents centres de compétences européens en matière de TAO : SYSTRAN, SITE, le laboratoire GETA et plusieurs équipes de recherche participant au projet EUROTRA. La participation de sociétés déjà utilisatrices de la TAO, comme RANK XEROX (UK) et KRUPP (RFA), donne au projet une dimension industrielle. Les différents partenaires, qui travaillaient jusqu'alors sur des projets concurrents, sont parvenus à la conclusion que la mise en commun des hommes, des techniques, des méthodes et des idées était le seul moyen d'aboutir à un système de TAO à la pointe de la technologie, et capable de faire face aux énormes efforts consentis par les Japonais dans ce domaine.

La terminologie et la lexicographie

En terminologie et lexicographie, SITE mène différentes activités concernant aussi bien la production de glossaires terminologiques et de lexiques, que la conception et la réalisation d'outils d'enrichissement, de consultation et de stockage de dictionnaires électroniques.

PHENIX, la base terminologique multilingue à l'usage des traducteurs

La base PHENIX, conçue et réalisée en interne, permet de gérer des fiches terminologiques constituées, pour chaque mot (ou expression), de sa(ses) traduction(s), de son contexte d'utilisation (domaine, sous-domaine, etc.), de quelques informations linguistiques (catégorie, genre, etc.), et d'un *code objet*. Ce dernier est en quelque sorte une caractérisation sémantique, véhiculant des informations précieuses pour un traducteur. On distingue environ 200 codes objets dans PHENIX (par exemple, *anomalie, métal, pays*).

Un système de gestion évolué permet aux traducteurs d'accéder aisément à cette base. On peut notamment exprimer des critères de sélection très raffinés pour trouver l'équivalent approprié dans un contexte donné (domaine, client, composant, etc.). L'accès est possible à partir de n'importe quelle langue (aucune langue n'est privilégiée comme source). Aujourd'hui, la base PHENIX comporte environ 240 000 entrées multilingues (français, anglais, allemand, espagnol) dans divers domaines techniques, en particulier l'aéronautique.

PHENIX fonctionne de plus en réseau, permettant ainsi le partage de l'expérience acquise à l'intérieur d'une équipe de traducteurs, qui l'enrichissent et la consultent. Pour assurer la qualité et l'homogénéité du contenu, deux terminologues spécialisés sont chargés de valider les nouveaux termes.

Remarquons que la terminologie, après avoir été introduite comme aide à la traduction, a acquis ses lettres de noblesse, puisqu'elle est vendue maintenant pour elle-même et non plus au titre d'une autre prestation. Cette prise de conscience de l'importance de la terminologie pour les entreprises ne se limite d'ailleurs pas à l'aspect traduction, mais englobe tous les aspects de la production de documents, et en particulier la rédaction.

BDTAO, la base lexicale dédiée à la TAO

Pour réduire les coûts des dictionnaires de TAO, tout en améliorant leur qualité, B'VITAL a mis au point une méthodologie d'enrichissement de ces dictionnaires. Le système BDTAO a été développé dans ce contexte pour permettre une gestion aisée et naturelle des données lexicales et la production des dictionnaires de TAO (système dédié à la TAO). Cette base sert à l'enrichissement et à la mise à jour des dictionnaires du système ARIANE F ⟶ A. BDTAO contient actuellement environ 20 000 termes français et anglais, à partir desquels sont produits automatiquement les différents dictionnaires (monolingues et bilingues) nécessaires au fonctionnement du système ARIANE. Les entrées de BDTAO sont très riches au niveau des informations morphologiques, syntaxiques et sémantiques.

Face au coût encore important de l'enrichissement manuel de BDTAO par les lexicographes, l'équipe a décidé de tenter une expérience de récupération automatique de données lexicales à partir de la base PHENIX : il s'agissait de charger automatiquement le glossaire du FALCON 900 dans BDTAO. Ce glossaire de 5000 entrées

bilingues, utiles à l'évaluation de ARIANNE F ⟶ A, a été chargé dans BDTAO avec un certain succès : environ 90 pour cent des termes ont été complètement et correctement convertis, notamment grâce à l'ajout d'un module de précalculs linguistiques dans le programme de conversion. Cette opération a permis d'économiser plusieurs mois-hommes par rapport à la méthode classique d'enrichissement manuel. La base PHENIX étant relativement riche (120 000 termes en français et en anglais), il devient donc possible de charger quasi automatiquement tous ces termes dans le système ARIANE F ⟶ A.

Un poste d'enrichissement lexicographique moderne

Le groupe SITE développe actuellement un poste d'enrichissement lexicographique et une base de données lexicales pour le CNET, dont les objectifs techniques sont proches de ceux affichés par les grands projets actuels du domaine, GENELEX et MULTILEX. Les développements sont en cours sur station de travail, et doivent aboutir à un système industriel fin 1991. Ce système est développé sous UNIX en respectant les standards actuels : C au niveau du langage de programmation, SQL au niveau de l'accès base de données, X-WINDOW/MOTIF pour ce qui est de l'interface homme-machine, SGML pour la définition du format d'échange des données, etc. À terme, un outil de ce genre sera utilisé comme base de référence pour le système de TAO.

Dans ce système, une entrée de la base lexicale est associée à des informations de natures très différentes : informations pour l'humain (définitions, exemples, etc.), informations pour le traitement automatique (transitivité, pronominalité, etc.). Ces informations sont modélisées sous forme d'attributs et de relations (relation d'abréviation, par exemple). Un réseau sémantique permettant de relier les concepts définis par les entrées lexicales est également mis à disposition du lexicographe.

Cette base est générique, dans la mesure où elle est indépendante d'une application ou d'une théorie linguistique particulière : l'utilisateur a les moyens de définir le type d'information qu'il désire associer à chaque terme, et de faire évoluer ces informations. Ainsi, la base n'est pas figée à un état donné de la technique. Cette caractéristique du système lui confère son indépendance vis-à-vis de la langue, et une évolution vers la prise en compte de données multilingues est envisagée.

Le poste de travail, ergonomique et productif, permet au lexicographe de saisir et de gérer toutes ses données. Ce poste met en œuvre toutes les techniques modernes de gestion et d'interfaçage (multi-fenêtrage, souris, boutons, icônes, etc.). Dans le but d'améliorer la productivité des lexicographes, le système met à disposition du lexicographe un L4G, lui permettant de définir :

- des accélérateurs linguistiques : il s'agit de précalculs linguistiques, fonction du lemme et de tous les champs déjà renseignés, qui permettent de préremplir un attribut ou de présélectionner un sous-ensemble de valeurs pertinentes dans une liste de valeurs possibles;

- des vérificateurs de cohérence : des fonctions permettant de vérifier différentes relations et dépendances inter-attributs seront activées soit en temps réel lors de l'enrichissement, soit en temps différé.

Cet outil générique permettra donc au CNET d'organiser, comme il l'entend, sa base de données lexicales. Il constitue en fait un point de départ très intéressant pour construire des dictionnaires électroniques répondant à différents besoins (terminologiques, lexicaux, dictionnaires du commerce, etc.). À partir de ces développements, SITE envisage de définir et de réaliser une version plus performante de PHENIX.

Vers un poste de travail intégré

Motivations et objectifs

Motivations

Les motivations pour un poste de travail moderne sont avant tout de type économique et technique. Il est clair que, avec la concurrence de plus en plus forte qui règne dans la plupart des grands domaines industriels (aéronautique, informatique, télécommunications, mécanique...), la qualité d'une documentation devient un facteur de décision important dans la prise de marché. Par exemple, la disponibilité de la documentation dans la langue du demandeur est un atout considérable.

Aujourd'hui, les coûts de constitution de la documentation sont souvent élevés, car celle-ci est souvent vue comme une activité annexe, mal intégrée au processus industriel, et toujours sous-évaluée (voire oubliée) lors de la définition d'un projet. Cet état des choses est heureusement en train de changer, et les industriels accordent de plus en plus d'importance à tous ces aspects : qualité de la rédaction et de la traduction, constitution d'une terminologie « normalisée », etc.

Des outils spécialisés puissants et adaptés

Dans le domaine de la production documentaire, il existe actuellement sur le marché un certain nombre d'outils spécialisés pour telle ou telle tâche. Citons, parmi les plus connus :

- les traitements de textes et les hypertextes,
- les outils dictionnaires : thésaurus, dictionnaires bilingues, dictionnaires de synonymes, etc.
- les outils de correction orthographique,
- les outils de gestion documentaire : indexation, interrogation (par mot clé ou autre),
- les outils de TAO : THAM (traduction humaine aidée par la machine) et TA (traduction automatique).

Notons que nombre de ces outils sont souvent difficiles à utiliser de manière simple dans les conditions réelles d'exploitation : par exemple, un rédacteur, travaillant sur un document technique français où il apparaît de nombreuses expressions anglaises (en informatique ou en aéronautique, par exemple), n'utilisera pas le correcteur orthographique français, qu'il jugera trop bavard et donc inefficace. De la même façon, un

outil de gestion documentaire ou un outil de TA, même de grande qualité, seront inutilisables s'il leur est impossible de gérer les ordres de mise en forme des documents.

Par ailleurs, de nouveaux outils commencent à apparaître (vérificateurs grammaticaux, vérificateurs stylistiques, analyseurs de corpus, etc.), qui étofferont la panoplie des utilisateurs. Les nouvelles possibilités qu'offre la technologie actuelle (processeurs rapides, réseaux distribués hétérogènes, bases réparties, écrans graphiques, CD-ROM, etc.) laissent penser que, dans 10 ans, le poste de travail aura subi des changements radicaux par rapport à son état actuel.

Une intégration poussée

La prolifération d'outils ne saurait être efficace que si l'ensemble de ces outils est intégré. Cela signifie que leur coexistence doit pouvoir être assurée sur un même poste de travail, et que la communication inter-outils est possible et transparente pour l'utilisateur final. En plus d'outils spécialisés tels que ceux présentés plus haut, le poste devra intégrer des outils plus classiques ou relevant d'autres techniques que celles qui nous intéressent ici : outils de DAO et de PAO, lecteur optique, serveur d'impression, lecteur de CD-ROM, accès à des bases de données textuelles et non textuelles (locales ou distantes), etc.

Il s'agit, en fait, de permettre à l'utilisateur d'effectuer toutes les opérations à partir du même environnement et de la façon la plus « économique » possible (sans refrappe...). Pour consulter un dictionnaire par exemple, il doit pouvoir sélectionner le terme dans le texte, puis activer la recherche. De même, un équivalent trouvé dans un dictionnaire doit pouvoir être inséré directement dans le fichier de travail (aussi simplement que s'il s'agissait d'un déplacement de texte).

C'est l'intégration de ces outils qui sera à la base des chaînes documentaires modernes, que les industriels visent pour la prochaine décennie. Cette démarche est analogue, et s'avère même complémentaire, à celle menée dans le cadre de l'initiative CALS du DoD, qui vise, dans un premier temps, à permettre la communication entre les diverses bases de données réalisées dans le cadre d'un projet industriel ou militaire (données de conception, de fabrication, de logistique, etc.), et, dans un second temps, à intégrer toutes ces données dans une base unique répartie (*publishing data base*), avec accès par vues multiples sur cette base.

Un poste multi-usage configurable

Le poste de travail (que nous avons jusqu'ici appelé *poste du rédacteur-traducteur*, pour simplifier nos propos) est, en fait, plus général que sa simple dénomination le laisse entendre (faut-il parler de *poste du communicateur*, comme certains le proposent?). Ce poste met en œuvre des outils qui sont communs à plusieurs métiers dans le domaine plus vaste de la gestion documentaire, et qui doivent procurer une aide pour différents types de tâches : rédaction, recherche terminologique, consultation et enrichissement de dictionnaires, traduction, pré et post-édition en TA, mise en page, etc.

Pour la définition du poste de travail, il faut éviter de tomber dans les pièges extrêmes qui sont :

- la définition d'un poste unique, utilisable par tous les métiers, mais qui sera trop riche et trop difficile à mettre en œuvre pour une utilisation particulière;
- la définition d'autant de postes qu'il y a de métiers, solution qui risque de conduire à une multiplication incontrôlable de configurations, étant donné l'évolution rapide des besoins et des métiers.

La définition d'un poste configurable qui offre à l'utilisateur toutes les fonctionnalités dont il a l'utilité, et celles-là seulement, apparaît comme la solution la plus ergonomique, et donc en définitive comme la plus productive. Le contrôle de l'activité sera laissé à l'utilisateur pour tout ce qui concerne les décisions de haut niveau : soumission de telle partie du texte à un système de TA, révision ou non du résultat, etc. Un tel poste permet de prendre en compte de manière souple les évolutions intervenant dans chaque métier, par adaptation de la configuration.

Mise en commun des compétences

Il est clair que la conception du poste de travail configurable nécessite non seulement l'intervention des divers métiers concernés (rédacteur, traducteur, terminologue, post-éditeur, etc.), mais également de spécialistes externes (ergonome, ingénieur en documentation, informaticien, spécialiste en bases de données, etc.). Remarquons qu'il ne s'agit ici que du développement du poste de travail, et non des divers outils qui y sont intégrés. Le coût de développement induit par cette entreprise est justifié par les enjeux que représente la mise au point d'une chaîne documentaire industrielle de bonne qualité.

Quelques fonctionnalités spécifiques au poste du traducteur

Pour terminer, nous décrivons un certain nombre de fonctionnalités, qui devraient être mises à la disposition du traducteur, utilisateur éventuel d'un système de TAO. Citons, par exemple :

- l'affichage simultané des textes source et traduit, avec des facilités de gestion synchronisée des deux fenêtres;
- les fonctions spécifiques : un traducteur, et en particulier un post-éditeur, appréciera certaines fonctions particulièrement adaptées à la manipulation textuelle (permutation intelligente de groupes de mots, rejet en fin de phrase, etc.). Ces fonctions auront une certaine connaissance de la structure textuelle;
- la manipulation grammaticales (mise au pluriel, changement du temps d'un paragraphe, mise à la voix passive/active, etc.);
- la recherche et le remplacement intelligents intégrant des connaissances grammaticales et typographiques;
- la détection des phrases répétitives, pour recherche automatique de leurs traductions;
- la gestion d'annotations successives (utiles pour les différentes relectures).

De manière générale, il s'agit d'introduire à tous les niveaux des fonctions intelligentes faisant appel à des processus linguistiques de différents types : lemmatiseurs (en particulier pour améliorer l'accès aux dictionnaires), analyseur grammaticaux (entre autres pour la vérification grammaticale), analyseurs stylistiques (pour la critique de style), etc. La disponibilité de corpus bilingues peut également permettre de réaliser, en amont des traducteurs et rédacteurs, des travaux qui soient réutilisables par eux : génération de dictionnaires bilingues, tables statistiques, etc. Par ailleurs, la communication avec d'autres spécialistes doit pouvoir se faire aisément : communication d'une liste de mots nouveaux au terminologue, questions au rédacteur, etc.

Conclusion

Vu l'évolution actuelle de la technologie, il n'est pas utopique d'envisager un poste de travail intégré et configurable, fournissant de nombreux outils d'aide à l'utilisateur. Toutes les activités présentées dans cet article sont en fait très liées, et l'expérience montre qu'elles doivent être considérées dans leur ensemble.

L'utilisation du TALN et des outils annexes dans les activités de gestion documentaire est devenue un axe stratégique pour de nombreuses entreprises. Pour SITE, la maîtrise d'œuvre du projet EUROLANG et le développement du poste d'enrichissement lexicographique pour le CNET s'inscrivent tout à fait dans cet axe.

Bibliographie

BACHUT D. (1991) « Industrialisation d'un système de TAO français-anglais dans le domaine technique », *Colloque Génie linguistique 91*, Versailles, 16-17 janvier 1991.

BOITET C., *et al.* (1982) « ARIANE-78 : an Integrated Environment for Automated Translation and Human Revision », *Coling 82*, Prague.

HEYER G. *et al.* (1991) « Motivation, Goals and Milestones of Esprit II project MULTIPLEX », *Colloque Génie linguistique 91*, Versailles, 16-17 janvier 1991.

MARET D. (1991) « Poste de travail lexicographique générique du CNET », proposé au *Colloque d'Avignon 91*.

NOSSIN M. (1991) « Le projet GENELEX : EUREKA pour les dictionnaires électroniques », *Colloque Génie linguistique 91*, Versailles, 16-17 janvier 1991.

VAUQUOIS B. et BOITET C. (1985) « Automated Translation at Grenoble University », *Computational Linguistics*, vol. 11, n° 1, January-March 1985.

PARTIE II

La station de travail du traducteur

Présidents : Philippe Thoiron
Jacques Goetschalkx

13

Pour le traducteur : un poste de travail à trois niveaux d'assistance

Alan MELBY

Brigham Young University at Provo, Provo, Utah, États-Unis

Déjà dix ans!

Quand on parle de l'état des choses dans dix ans, il est utile de regarder le progrès fait depuis dix ans. En 1981, l'auteur a conçu un poste (ou une station) de travail pour traducteur. Il a présenté les principes de ce poste au Congrès international de linguistique computationnelle (COLING 82).

L'auteur croyait (en 1981) qu'en moins de cinq ans un PTT (poste de travail du traducteur) se trouverait sur à peu près chaque bureau où un traducteur exerce son métier. Certes, beaucoup de traducteurs ont aujourd'hui un micro-ordinateur individuel, et emploient un système de traitement de texte. Presque tous les jeunes traducteurs ont l'aptitude voulue pour taper et corriger un texte. Et plusieurs emploient la télécommunication au-delà de l'envoi de documents par télécopieur. Mais ce seront les prochaines dix années qui verront l'emploi massif et intégré des outils informatiques qui entourent le traitement de texte. Parmi ces outils, qui existaient tous il y a dix ans, se trouvent les suivants :

niveau I
- des aides à la rédaction,
- la gestion des données terminologiques;

niveau II
- l'analyse du texte source,

- la consultation automatique des fichiers de terminologie,
- la synchronisation de textes source et cible;

niveau III

- l'accès à la traduction automatique.

La TA et le PTT

Un mot sur la TA (traduction automatique) et le poste du traducteur. Il y a dix ans, plusieurs prévoyaient les progrès rapides de la traduction automatique. Aujourd'hui, seulement un petit pourcentage de ce qui est traduit passe par la TA. Des études récentes en linguistique, loin de suggérer que les traducteurs devraient se faire du souci pour la décennie à venir, nous aident à comprendre pourquoi les systèmes de TA actuels ne pourront jamais être utilisés pour la plupart des textes généraux (voir Melby, 1991a).

L'échange de données terminologiques

Un autre aspect très important du poste de travail du traducteur, qui ne peut se faire par un individu, est l'accord sur un format d'échange pour les fichiers de terminologie.

Il est bien connu que les traducteurs ont toujours eu tendance à garder leurs fiches terminologiques pour eux-mêmes. Néanmoins, ceci devrait changer, pour plusieurs raisons. En voici trois :

1. Les organisations qui créent des documents vont exiger une meilleure coopération tout au long de la chaîne de production, c'est-à-dire entre les terminologues, les auteurs, les traducteurs, les réviseurs et les imprimeurs afin d'augmenter la qualité des textes et d'éviter le travail inutile.
   Le texte d'un document sera transmis sous forme électronique pour ne pas le saisir à nouveau, et un fichier de terminologie sera transmis avec le document pour empêcher des changements terminologiques. On peut vérifier à chaque étape si la terminologie du document est conforme au standard.
2. Dans une organisation à plusieurs départements, une coordination pour la terminologie est nécessaire afin d'améliorer la communication interne et externe. Il est fâcheux pour un client de recevoir deux documents de la même société dans lesquels des termes différents sont employés pour désigner la même réalité.
3. Les bureaux de standards (d'une industrie, du gouvernement, et internationaux) vont de plus en plus chercher de nouveaux moyens pour distribuer les données terminologiques et pour mieux alimenter leurs bases de données.

Mais, évidemment, nous n'aurons jamais tous le même matériel ou le même progiciel de gestion terminologique. Bien sûr, on peut imprimer un fichier et échanger sur papier, mais pour profiter de la rapidité d'accès et du traitement automatique (y-compris l'insertion d'un terme dans un document sans le taper) de fichiers termi-

nologiques, il faudra un format d'échange très flexible et indépendant d'un matériel ou d'un logiciel particulier.

L'auteur se concentre, actuellement, sur la définition d'un tel format en dirigeant un comité de TEI (*Text Encoding Initiative*) sur la terminologie. Ce comité (A&I-7) est chargé de définir un type de document pour échanger des données terminologiques entre des progiciels différents et divers matériels. Ces documents seront conformes à la norme SGML.

SGML est un standard international (ISO 8879) qui a paru en 1986 et qui reçoit aujourd'hui beaucoup d'attention. On peut penser à SGML comme langage de programmation déclaratif (en anglais — *declarative vs. procedural*) dans lequel on écrit des descriptions de types de documents. Une telle description formelle s'appelle un DTD. Le TEI (soutenu, entre autres, par l'*Association for Computational Linguistics*) a pour but l'écriture de plusieurs DTD qui vont devenir, on l'espère, des standards.

Le comité A&I-7 a le devoir de définir un DTD pour l'échange de données terminologiques, et les résultats seront publiés dans le livre *Guidelines for Encoding and Interchange of Machine-Readable Texts* du TEI.

Beaucoup d'efforts restent à faire, mais l'avenir est très prometteur pour le poste de travail du traducteur.

Bibliographie

MELBY A. (1991a) « Des causes et des effets de l'asymétrie partielle des réseaux sémantiques liés aux langues naturelles », *Cahiers de lexicologie*, n° 58, 1991-1, Paris, Didier Érudition, p. 6-43.

MELBY A. (1991b) « Micromater. A Proposed Standard Format for Exchanging Lexical/ Terminological Logical Data File », *Meta*, vol. 36, n° 1, p. 135-160.

14

Le poste de travail de traducteur à l'ONU — Horizon 2001

Michel GAINET

Organisation des Nations Unies, New York, États-Unis

Introduction

L'Organisation des Nations Unies, comme toute organisation qui emploie de très nombreux traducteurs — elle en compte quelque 600 répartis entre ses trois centres principaux (New York, Genève et Vienne) et de nombreux autres bureaux (Addis-Abeba, Bangkok, Nairobi, Bagdad, Santiago du Chili, etc.) — s'intéresse depuis de nombreuses années aux innovations technologiques susceptibles d'améliorer qualitativement et quantitativement la fonction « traduction ». C'est ainsi que s'est constitué il y a trois ans au sein de la Division de traduction un groupe de travail des innovations technologiques qui contribue à définir la politique de la Division en la matière et suit les progrès accomplis dans des domaines allant du traitement de texte à la traduction automatique.

Pourtant, l'ONU est sans doute en retard sur la plupart des autres organisations internationales pour ce qui est de l'équipement de ses traducteurs. Si l'on fait abstraction de quelques machines qui servent à tester des applications présentant un intérêt potentiel pour les traducteurs, le poste de travail du traducteur onusien est, en 1991, à peu près ce qu'il était il y a 30 ans. Cet état de choses tient certes à diverses contraintes, qui résultent notamment de la structure même de l'Organisation et de l'austérité budgétaire à laquelle nous sommes astreints depuis des années, qui fait que nos organes de tutelle financière hésitent à approuver des projets dont il est difficile de chiffrer par avance

les gains de productivité qu'ils permettront. Mais il tient aussi à l'approche adoptée par la Division de traduction et son Groupe de travail des innovations technologiques.

L'approche de la Division de traduction

La Division a en effet pris le parti d'attendre, pour équiper ses traducteurs de postes de travail informatisés, qu'il existe un ensemble d'outils électroniques qui contribuent véritablement à améliorer leurs conditions de travail. C'est ainsi que l'ONU, à la différence de nombreuses autres organisations, s'est refusée à équiper ses linguistes de micro-ordinateurs qui auraient servi d'abord au traitement de texte; d'une part, parce qu'elle considérait qu'il serait néfaste, psychologiquement, de donner aux traducteurs — qui, dans leur grande majorité, ont l'habitude de dicter leurs textes — l'impression que l'informatisation avait pour but premier de permettre des économies en aval de la traduction; d'autre part, parce que des applications susceptibles d'avoir un retentissement considérable sur le travail du traducteur — dans le domaine de la terminologie, mais surtout dans celui de la gestion de la documentation de référence — étaient en cours de développement. En équipant hâtivement les traducteurs, l'ONU courait le risque d'acquérir du matériel et des logiciels qui se seraient à brève échéance révélés inadaptés aux exigences techniques de ces applications.

Puisque la base de matériel installé est encore extrêmement modeste, il serait facile de céder à la tentation du futurisme, et de décrire pour l'an 2001 un poste de travail mettant en œuvre toutes les technologies qui sont actuellement en plein essor : traduction automatisée, saisie des originaux-papiers au scanner avec reconnaissance optique des caractères, dictée directe à une machine à traduire capable de reconnaître la parole, etc. Notre groupe de travail suit certes très attentivement toutes ces innovations, mais je me bornerai ici, plus modestement, à esquisser les caractéristiques d'un poste de travail de traducteur adapté aux conditions propres à l'ONU, conditions qui tiennent à la fois à la structure de l'Organisation et de ses services linguistiques et à la nature des travaux confiés aux services de traduction. Je ne m'étendrai pas sur les caractéristiques qui sont plus ou moins communes à tous les postes de travail de traducteur en l'état actuel de la technique, mais j'insisterai par contre sur une application — la gestion de la documentation de référence grâce à une banque de textes sur disques optiques — pour laquelle l'ONU joue à certains égards un rôle de pionnier.

Contraintes organisationnelles, nature des travaux et caractéristiques du poste de travail de traducteur

a) L'ONU est une organisation largement décentralisée, et possède des services de traduction non seulement dans ses principaux centres — Siège de New York, Office des Nations Unies à Genève et Office des Nations Unies à Vienne — mais aussi dans de nombreux bureaux répartis dans le monde entier (commissions économiques et sociales pour l'Afrique, l'Amérique latine et les Caraïbes, l'Asie et le Pacifique, l'Asie occidentale, Programme des Nations Unies pour l'environnement et Centre

HABITAT, à Nairobi, etc.). Cette décentralisation exige que les postes de travail de traducteur mettent en œuvre du matériel et des logiciels standardisés, pour que le personnel linguistique, dont l'ONU encourage la mobilité, n'ait pas à suivre un nouvel apprentissage à chaque réaffectation. Cette décentralisation exige aussi que les traducteurs, où qu'ils travaillent, aient facilement accès à la documentation produite dans n'importe quel autre centre. C'est notamment en fonction de cette exigence qu'a été conçu le système à disques optiques dont je parlerai plus en détail tout à l'heure.

b) Les services auxiliaires situés en amont de la traduction (recherche documentaire et surtout terminologie) ont des effectifs modestes par rapport à ceux des linguistes. Il s'ensuit que les services de traduction de l'ONU participent davantage aux travaux terminologiques que s'ils disposaient d'un effectif de terminologues plus nombreux, et que leur poste de travail doit être conçu pour les besoins de ces apports terminologiques, c'est-à-dire permettre la saisie et la recherche de données terminologiques dans les six langues officielles, et être raccordé à une base de données multilingue à laquelle seront intégrés leurs apports, une fois ceux-ci validés par des terminologues professionnels. Ce type d'application ne soulève guère de difficultés s'il se limite à des langues utilisant l'alphabet latin, mais pose des problèmes ardus lorsqu'il s'agit d'intégrer dans une base de données accessible depuis et vers toutes les langues officielles le russe, l'arabe et surtout le chinois. J'ajouterai que l'ONU a l'obligation politique d'assurer l'égalité de traitement de toutes les langues officielles, et que cette règle s'applique à la diffusion, sous forme électronique ou sous forme imprimée, de données terminologiques. Un gros travail de développement est en cours pour la mise en place d'une base de données tournant sur micro-ordinateurs et remplissant ce difficile cahier des charges.

c) La nature des travaux confiés aux services de traduction dicte aussi nombre des caractéristiques d'un poste de travail de traducteur.

L'ONU, n'étant pas une institution spécialisée, produit des documents relevant de domaines extrêmement divers — textes politiques et économiques, bien sûr, mais aussi documents traitant de domaines de pointe tels que la réglementation des activités spatiales ou de l'exploitation minière des grands fonds marins, le désarmement, etc. Cette diversité et le faible volume de la documentation à forte densité de termes techniques ont dissuadé jusqu'à présent l'ONU de se lancer dans la traduction automatique, des essais effectués ces dernières années avec la plupart des systèmes disponibles sur le marché ayant donné des résultats inacceptables.

En revanche, cette même diversité exige que le poste de travail du traducteur permette à celui-ci d'accéder à des banques de données conçues spécifiquement ou non pour les linguistes, à l'intérieur de l'Organisation comme à l'extérieur. Le poste de travail devra donc être équipé pour les communications avec ces banques de données et aussi être raccordé à des lecteurs de CD-ROM accessibles en réseau.

d) Une forte proportion des documents de l'ONU font de larges emprunts, explicites ou non, à des documents antérieurs. D'autres, quoique entièrement nouveaux, ne sont compréhensibles que si on les rapproche d'autres documents antérieurs ou parallèles. Pour les traducteurs, la fonction de recherche documentaire revêt donc

une extrême importance : dans certains cas (résolutions, conventions, certains textes politiques notamment), il est impératif de reprendre exactement les termes des documents de référence; dans d'autres, la recherche des « précédents » fait simplement économiser un temps considérable au stade de la traduction. Pour se faire une idée de l'importance de la fonction de recherche documentaire il suffit de savoir que rien qu'à New York une trentaine de documentalistes s'y consacrent, constituant, pour chaque document soumis et pour chacune des langues de traduction, un dossier qui renferme les textes cités ou plus ou moins repris. Leur travail est rendu difficile, entre autres choses, par la décentralisation de l'Organisation, les documents produits à Vienne, par exemple, n'étant dans le meilleur des cas disponibles à New York que deux ou trois semaines après leur parution. De plus, la recherche de « citations cachées » exige une longue habitude des diverses catégories de textes, et le traducteur est souvent appelé à compléter le travail fait par les documentalistes.

Aussi, le Groupe de travail de la Division de traduction s'est-il préoccupé dès sa création de rationaliser et de rendre encore plus efficace cette fonction. Lorsque l'ONU a décidé, fin 1989, de se doter d'un système à disques optiques pour l'archivage et la consultation des documents parlementaires, le Groupe a vu le parti qui pouvait être tiré d'un tel système aux fins de la recherche documentaire, et a fait en sorte que le cahier des charges tienne compte des besoins des traducteurs.

Ce système sera mis en place d'ici quelques mois. Installé initialement à New York et à Genève, il sera par la suite étendu aux autres centres des Nations Unies.

Il permettra la saisie en temps réel de tous les documents « parlementaires » produits à New York et à Genève, sur des disques optiques non réinscriptibles de grande capacité (6,4 GO). La saisie était initialement prévue en mode image seulement (saisie des documents-papier au scanner), mais, à la demande essentiellement des services linguistiques, il est maintenant prévu de saisir les documents en code (mode caractères) ce qui non seulement réduira considérablement la capacité de stockage requise pour un nombre donné de pages, mais aussi, et surtout, permettra aux usagers — notamment aux traducteurs — d'effectuer des recherches à l'intérieur des documents extraits du système, par exemple pour faire ressortir les différences entre deux versions successsives d'un rapport, et d'intégrer directement à une traduction en cours les passages repris de documents antérieurs.

Les documents seront indexés selon les principaux descripteurs du Système d'information bibliographique de l'ONU (SIBONU/UNBIS), toujours en temps réel, de sorte que la version électronique de tous les textes parlementaires sera disponible dès la parution de la version imprimée. D'autre part, un échange quotidien de documents entre New York et Genève permettra de disposer, dans chacun des deux centres, d'une collection complète des documents produits de part et d'autre de l'Atlantique.

La consultation des documents se fera à partir de micro-ordinateurs standard équipés d'une interface graphique spéciale et d'un moniteur haute définition. Des documents extraits du système pourront être imprimés localement sur une imprimante laser et se présenteront alors exactement comme la version papier originale, avec en-têtes

et graphiques. La téléconsultation sera possible grâce à des interfaces compatibles avec les réseaux téléphoniques commutés et avec les nouveaux réseaux numériques à intégration de services (RNIS), ce qui sera fort utile non seulement à des usagers tels que les missions des États membres auprès de l'ONU ou des bibliothèques, mais aussi aux traducteurs en poste dans des bureaux décentralisés, qui actuellement ont beaucoup de difficulté à obtenir la documentation de référence dont ils ont besoin.

Telles sont les principales caractéristiques qui feront l'originalité du poste de travail de traducteur à l'ONU. Je ne dirai rien des caractéristiques ergonomiques — fenêtrage, taille et définition des moniteurs, etc. — car les services linguistiques de l'ONU ne se sont pas encore penchés sur ces aspects. J'ajouterai simplement que le matériel mis à la disposition des traducteurs devra être suffisamment puissant non seulement pour fonctionner efficacement en mode multi-tâche, mais aussi pour offrir une marge suffisante en prévision d'applications futures.

Conclusion

En guise de conclusion, je me permettrai de me livrer à quelques spéculations sur la manière dont nos traducteurs réagiront à leur nouvel environnement de travail. Je constate, et pas seulement chez les jeunes traducteurs, que face aux innovations technologiques, l'impatience d'essayer l'emporte largement sur la peur de la machine. Un certain nombre de traducteurs, déçus de la lenteur que met l'Organisation à les équiper, n'ont pas hésité à investir eux-mêmes dans du matériel et des logiciels. J'espère qu'étant donné le parti que nous avons choisi, le nouvel environnement traductionnel sera perçu comme propre à libérer le traducteur des aspects les plus ingrats de sa tâche et à lui permettre de se concentrer sur les aspects les plus gratifiants — et aussi les plus difficiles — pour lesquels la concurrence de la machine ne semble pas à craindre de sitôt.

15

La technologie au service du traducteur

Klaire TREMBLAY

Secrétariat d'État du Canada, Langues officielles et traduction, Ottawa, Canada

La Loi canadienne sur les langues officielles stipule que tout citoyen canadien a le droit de recevoir les services du gouvernement fédéral dans la langue officielle de son choix, soit en anglais, soit en français. C'est donc dire que tous les ministères et toutes les agences gouvernementales doivent offrir leurs services dans ces deux langues. Dans un tel contexte, la traduction joue un rôle primordial. La responsabilité de fournir les services nécessaires a été confiée au Secrétariat d'État du Canada, plus particulièrement au Bureau de la traduction. Vous vous doutez bien que la tâche est énorme et la demande de traduction, astronomique. À titre d'exemple, les 800 traducteurs du Bureau de la traduction ont traduit plus de 297 millions de mots l'an dernier.

Soumis à une telle pression de la demande, le Bureau de la traduction est constamment à la recherche de moyens d'accélérer ou, du moins, de faciliter le travail de ses traducteurs. La nouvelle technologie offre toute une panoplie d'aides à la traduction. L'intérêt du Bureau pour ce champ d'activité ne date toutefois pas d'hier. En effet, les premiers travaux de recherche en matière de technologie appliquée aux industries de la langue ont commencé dès le milieu des années soixante, en collaboration avec l'Université de Montréal. Ce partenariat a permis de mettre au point le système de traduction automatique TAUM-MÉTÉO. Ce système traduit aujourd'hui tous les bulletins météorologiques du ministère canadien de l'Environnement. Le processus s'effectue presque automatiquement et l'intervention humaine est minimale. Pour vous donner une idée de l'économie de temps réalisée grâce à ce système, un traducteur met en moyenne 3 minutes 8 secondes à traiter, ou à réviser si vous préférez, un bulletin météorologique. La traduction de ce même bulletin lui aurait normalement pris de 30 à 40 minutes.

Le Bureau est également responsable du volet terminologique. Il possède l'une des plus importantes banques de terminologie du monde. Je parle évidemment de TERMIUM. La banque compte plus de un million de termes. Elle est maintenant accessible non seulement en ligne, mais également sur disque optique, ou CD-ROM.

Comme je viens de le mentionner, le Bureau de la traduction est constamment à la recherche d'outils qui lui permettront de répondre à une demande sans cesse croissante. Cette recherche s'est poursuivie pendant les années quatre-vingt. L'essai opérationnel de deux systèmes de traduction s'est révélé décevant et nous a amenés à faire preuve de la plus grande prudence. Les progrès dans le domaine de l'intelligence artificielle nous ont cependant incités à tenter une nouvelle expérience en traduction assistée par ordinateur. Je parle ici de l'essai du logiciel Logos.

Ce projet-pilote lancé en 1987 se déroule dans six services de traduction. En tout, plus de trente traducteurs et personnes de soutien mettent le logiciel Logos à l'épreuve chaque jour dans des champs d'application différents. Le rôle du traducteur consiste à évaluer les valeurs ergonomique et linguistique de l'outil de travail et à en déterminer l'utilité.

Le traducteur qui utilise Logos intervient uniquement au moment de l'enrichissement du dictionnaire et de la post-édition. La traduction se déroule de façon purement automatique, en arrière-plan. Au stade de la post-édition, les fonctions du traducteur s'assimilent davantage à celles du réviseur traditionnel. Le traducteur qui utilise Logos doit modifier considérablement ses méthodes de travail et faire face à un changement radical de son milieu.

Il va de soi qu'un tel projet nécessite une étroite collaboration entre fournisseur et utilisateurs. C'est la raison pour laquelle au cours des quatre dernières années, nous avons soumis à la société Logos plusieurs centaines de recommandations afin d'améliorer le système et de le façonner davantage en fonction de nos besoins. Malgré ces rapports directs avec la société Logos, force nous fut de constater au fil des ans que les logiciels de traduction disponibles sur le marché ne pouvaient répondre à nos besoins généraux. Il fallait trouver un autre outil plus polyvalent et plus convivial. Plutôt que de nous orienter vers un autre automate de traduction, nous avons opté pour un outil capable de faciliter ou d'informatiser certaines tâches connexes à l'acte traductionnel.

C'est ainsi qu'en 1989, le Bureau de la traduction a mandaté le Centre canadien de recherche sur l'informatisation du travail (CCRIT) pour la conception d'un poste de travail du traducteur très perfectionné. Le poste ne devait exploiter que du matériel et des logiciels courants. Le premier prototype a été mis en place dans trois services de traduction. Le poste comportait le matériel suivant : un micro-ordinateur AT 286 avec écran pleine page monochrome, une imprimante matricielle et une souris. Pour leur part, les aides informatisées regroupaient un traitement de texte à fonctions augmentées comprenant notamment des vérificateurs orthographiques anglais et français; un logiciel de comparaison de textes, qui décèle les passages ajoutés, supprimés ou modifiés; un outil de recherche documentaire, qui relève entre autres les occurrences de différents termes; un conjugueur électronique semblable au Bescherelle; un logiciel de gestion de données terminologiques, qui permet notamment la création et la fusion de glossaires maison; un outil de conversion de fichiers de traitement de texte; deux

dictionnaires électroniques bilingues; et, pour gérer efficacement et simultanément toutes ces applications en mémoire centrale, un logiciel de fenêtrage. Chaque poste de travail était également doté d'un logiciel de sauvegarde, d'un détecteur de virus et d'un mobilier ergonomique.

Pendant un an, douze traducteurs ont mis à l'essai le poste de travail en milieu opérationnel. Au cours de cette période, nous nous sommes entretenus régulièrement avec les traducteurs pour connaître l'utilité, l'efficacité et la convivialité de chacune des composantes. Il fallait également déterminer la valeur ergonomique de l'outil dans son ensemble. À l'issue de cette période d'essai, nous avons rédigé un rapport d'évaluation de la convivialité du PTT. Ce rapport regroupe les commentaires et les recommandations formulés par les traducteurs. Il présente également des recommandations en vue de l'amélioration des prototypes à venir.

La première année de l'expérience nous a permis de faire des constatations fort intéressantes sur les plans informatique et ergonomique et de rectifier notre tir avec le deuxième prototype. Je vais laisser à M. Elliott Macklovitch, du CCRIT, le soin de vous parler des problèmes d'interface survenus et des solutions apportées. Je vais plutôt m'attarder aux volets opérationnel et ergonomique du point de vue du traducteur.

Dans un premier temps, il est intéressant de constater que l'expérience vécue par les traducteurs diffère sensiblement d'un service de traduction à l'autre. En effet, alors qu'une équipe disposait d'assez de temps pour se familiariser avec les diverses fonctionnalités du nouvel outil de travail et pour constituer des dictionnaires spécialisés, une autre équipe devait souvent se limiter au traitement de texte en raison des impératifs du service, c'est-à-dire des textes urgents. Pour vous donner une idée, pendant la première année, l'une des équipes a créé quelque 3000 fiches terminologiques. Elle a en plus fusionné d'autres glossaires dans un seul et même dictionnaire qui compte maintenant plus de 24 000 termes. La deuxième équipe, elle, a créé au total une quarantaine de fiches.

Dans un deuxième temps, nous nous sommes rendu compte que l'intégration d'un poste de travail perfectionné entraîne de nombreuses difficultés techniques qui perturbent le travail des traducteurs. Il convient toutefois de souligner que, dans bien des cas, le manque de connaissances préalables du micro-ordinateur de la part du traducteur est à la source des problèmes. Il est donc indispensable pour un service de traduction doté de postes de travail perfectionnés d'avoir recours sur place aux services d'un « expert technique » en cas de panne. En effet, rares sont les traducteurs qui possèdent les connaissances informatiques voulues pour assurer le dépannage. Les interruptions sont donc nombreuses et irritantes pour les traducteurs.

Sur le plan organisationnel, la baisse de productivité attribuable au processus d'apprentissage provoque souvent des tensions entre les participants au projet-pilote et leurs collègues du service. En effet, ces derniers considèrent qu'ils font les frais de l'implantation de cette nouvelle technologie en absorbant la charge de travail des traducteurs sur PTT. Remarquez que le même phénomène se retrouve dans les sites Logos. Les traducteurs du service estiment que ceux travaillant avec Logos bénéficient d'un traitement de faveur du fait qu'ils ne participent pas aux urgences.

C'est au niveau de la formation et du temps consacré à l'apprentissage des fonctionnalités de l'outil que nous avons tiré la plus grande leçon. Premièrement, avant même de procéder à la formation comme telle, il est essentiel que les traducteurs maîtrisent très bien le traitement de texte et qu'ils possèdent de bonnes connaissances de base du micro-ordinateur. Il faut ensuite adopter une formule individualisée comportant des exercices pratiques. Il faut surtout procéder par étapes, c'est-à-dire enseigner l'usage d'un ou de deux logiciels à la fois. En effet, il faut éviter d'adopter une méthode de formation globale; les traducteurs sont incapables d'assimiler toutes les fonctionnalités du poste en même temps et ils ont tendance par la suite à sous-utiliser des fonctions importantes. Lors de l'implantation d'un poste de travail de technique de pointe, il ne faut surtout pas commettre l'erreur de sous-estimer la complexité de l'apprentissage à réaliser, en particulier pour les personnes qui ne sont pas familières avec les ordinateurs. N'oubliez pas que certains traducteurs doivent passer du dictaphone au PTT alors qu'ils n'ont jamais posé les doigts sur un clavier! Nous recommandons par conséquent une démarche plus graduelle. Outre la formation, la période d'apprentissage revêt une importance capitale. Les traducteurs doivent absolument disposer de suffisamment de temps pour explorer les possibilités du poste et pour approfondir les connaissances fraîchement acquises. Il est même souhaitable de réduire les objectifs de production des traducteurs sur PTT et de modifier leur charge de travail pendant les premiers mois.

Nous avons également appris quelque chose sur le plan ergonomique. Il est essentiel de prévoir suffisamment d'espace de bureau pour permettre un aménagement fonctionnel du poste de travail. L'encombrement des aires de travail rend difficile l'utilisation efficace du matériel. Les traducteurs sont souvent forcés de mettre au point des solutions artisanales aux problèmes de fonctionnement dans des espaces réduits. Dans un de nos services par exemple, les traducteurs ont dû improviser un système de cordes et d'épingles à linge pour atteindre les divers formats de papier continu derrière les imprimantes!

Pour ce qui est de l'ergonomie du poste lui-même, les traducteurs n'aiment pas l'écran pleine page à cause de la petite dimension des caractères en format de 66 lignes. De plus, la hauteur de l'écran pose des problèmes aux personnes de petite taille et à celles qui portent des verres à dioptrie progressive ou avec foyers. Quant à l'imprimante matricielle, les traducteurs l'estiment insatisfaisante en raison des fréquents blocages du papier continu et de la lenteur d'impression. Ils la trouvent aussi trop bruyante dans un environnement de travail à aires ouvertes.

Jusqu'à maintenant, les traducteurs sont satisfaits des diverses aides du poste. Leur préférence porte toutefois sur les outils de vérification orthographique et grammaticale et sur le logiciel de gestion de données terminologiques. Pour ce qui est de l'outil de fenêtrage dont l'utilisation peut sembler complexe, il est intéressant de constater que le degré de satisfaction est largement tributaire du nombre de logiciels utilisés. En effet, un traducteur qui ne maîtrise que le traitement de texte a forcément de la difficulté à intégrer l'aspect multidimensionnel du PTT à ses pratiques de travail. D'autres outils se sont révélés plutôt inutiles, tels les logiciels de recherche documentaire et de conversion de fichiers de traitement de texte.

Compte tenu des recommandations formulées par les traducteurs, nous avons amélioré le poste de travail en attendant la mise en place, en 1992, de la version réseau. Les améliorations ont été apportées surtout au niveau ergonomique. Le micro 286 a été remplacé par un modèle 386 SX plus performant, et l'écran pleine page a cédé la place à un écran VGA couleur. En outre, les traducteurs disposent chacun d'un lecteur de disque compact leur donnant l'accès direct à TERMIUM sur CD-ROM. Un mobilier ergonomique mieux adapté a été installé. De plus, tous les logiciels ont été mis à niveau et les deux outils inutiles ont été supprimés.

Dans l'ensemble, les traducteurs sur poste de travail estiment que le PTT a eu un effet bénéfique sur leur travail. Il a enrichi la tâche de traduction par une plus grande autonomie et une plus grande souplesse d'exécution. Il faut toutefois bien comprendre que le PTT ne permet pas de traduire plus rapidement, car il n'automatise pas l'acte traductionnel comme tel. Même avec l'ordinateur le plus performant, le traducteur ne peut penser plus vite!

Outre les deux grands projets Logos et PTT dont je viens de vous parler, le Bureau de la traduction poursuit ses efforts de recherche dans le domaine de la terminologie. Il procède actuellement à la mise au point d'une station de travail pour terminologue appelée LATTER dont le premier prototype devrait voir le jour d'ici un an. Cette station permettra notamment aux terminologues de traiter toutes les données terminologiques qui leur seront transmises de façon électronique par les traducteurs selon des paramètres préétablis.

Quand on parle d'outils de haute technologie, il est un élément extrêmement important qu'il ne faut surtout pas négliger, l'utilisateur. Depuis les débuts de nos projets de pointe, nous avons toujours insisté sur le rôle primordial du traducteur dans le développement. La plupart des traducteurs affectés à ces projets sont conscients du caractère indispensable de leur participation. Ils désirent que l'on reconnaisse leur apport essentiel au développement de puissants outils de travail axés sur l'avenir. Il ne serait pas réaliste de compter uniquement sur la satisfaction que ces professionnels de la traduction retirent de leur participation à ces projets. Il ne faut pas oublier que le recours aux micro-ordinateurs, aux aides à la traduction et aux logiciels de TAO exige des traducteurs un changement radical de leurs habitudes de travail. Les nouvelles aptitudes qu'ils doivent acquérir et la transformation complète de leur milieu de travail justifient à leurs yeux une reconnaissance spéciale. De son côté, la haute direction du Bureau de la traduction estime qu'il est impératif que les traducteurs participent au développement de toute nouvelle technologie susceptible d'avoir une incidence sur leur profession. Par conséquent, le Bureau s'efforce de maintenir ou d'améliorer le niveau de motivation de ces traducteurs par divers moyens, notamment en organisant des rencontres d'utilisateurs et des ateliers réunissant concepteurs de logiciels et traducteurs.

Mais quel sera l'outil du traducteur de l'an 2001 au Bureau de la traduction? Il y a longtemps que nous nous sommes posé cette question, et des plans sont déjà sur la table à dessin. Dès que nous aurons trouvé la solution réseau qui répondra à l'ensemble de nos besoins, nous intégrerons en un seul et même système les meilleures aides à la traduction disponibles sur le marché. C'est ainsi que le traducteur de demain pourra, de son poste de travail, communiquer directement avec son client, consulter toutes les

sources de documentation et les bases de données terminologiques électroniques nécessaires, confier certains passages de son texte à un logiciel de traduction, que ce soit Logos ou un autre, transmettre directement le résultat de ses recherches terminologiques à TERMIUM et, le cas échéant, procéder à l'éditique de son texte. Le tout intégré sur un seul micro-ordinateur. Les possibilités sont immenses dans le domaine de la bureautique, et le Secrétariat d'État du Canada déploie des efforts soutenus afin de demeurer à la tête du peloton.

Annexe I

Le poste de travail du traducteur (PTT)

Contexte

Avec l'implantation progressive des technologies modernes sont apparus de nouveaux besoins chez les traducteurs. Qu'il s'agisse de traductique, de bureautique ou d'éditique, le Bureau de la traduction s'emploie depuis des années à doter les traducteurs d'outils de pointe visant à informatiser certaines tâches dites connexes à l'acte traductionnel même.

C'est dans cet ordre d'idées que le Centre canadien de recherche sur l'informatisation du travail (CCRIT) s'est vu confier le mandat de concevoir un poste de travail pour traducteur, en collaboration avec la Direction de l'informatique. Le premier prototype a été mis à l'essai dans trois services de traduction en octobre 1989. Il a maintenant cédé la place au PTT 1.1, mieux adapté et plus performant.

Prototype

Le prototype repose sur un concept général, soit l'exploitation de matériel et de logiciels courants permettant d'informatiser, ou d'accélérer, certaines fonctions liées à l'acte de traduction, ce dernier demeurant le fait du traducteur lui-même.

La coquille de l'outil consiste en un micro-ordinateur de type IBM compatible 386 SX, muni d'une carte d'extension de mémoire INTEL AboveBoard Plus de 2 Mo, d'un écran couleur, d'une souris et d'une imprimante laser. Le noyau, lui, regroupe un ensemble de logiciels choisis parmi la panoplie d'aides informatiques fonctionnant en MS-DOS et offertes sur le marché. Ces logiciels ont tous fait l'objet de tests en laboratoire visant à déterminer leur convivialité et leur performance à l'égard des besoins exprimés par les traducteurs.

Le poste repose sur le logiciel de traitement de texte WordPerfect 5.1, devenu en quelque sorte la norme au sein du gouvernement fédéral. Viennent se greffer ensuite le logiciel de conjugaison Verbatex, les dictionnaires bilingues anglais-français Harrap's et Robert & Collins, et le vérificateur orthographique Chandioux, lesquels se révèlent d'une grande utilité lors de la préparation du texte et au moment de la relecture. Le traducteur dispose également du logiciel de comparaison de textes CompareRite, qui décèle les passage ajoutés ou supprimés; du logiciel Software Bridge, outil de conversion de fichiers de traitement de texte; du logiciel de gestion de données terminologiques Termex, qui permet notamment la consultation et la mise à jour de lexiques personnels ou communs; de l'outil de recherche documentaire Textsearch, qui relève entre autres les occurences de différents termes; et, finalement, du programme de sauvegarde

Fastback. Pour gérer efficacement et simultanément toutes ces applications en mémoire centrale, le logiciel de fenêtrage très performant DESQview a été intégré au poste de travail. Chaque traducteur dispose également de la dernière version du disque optique (CD-ROM) de TERMIUM.

Essai en milieu opérationnel

Depuis octobre 1989, trois services de traduction servent de banc d'essai, soit la DSTM Finances/ Conseil du Trésor à Ottawa, la section Économique et juridique de Montréal et la DSTM Communication à Ottawa. Les utilisateurs, qui, pour la plupart, sont des néophytes de l'informatique, ont pour tâche de vérifier la convivialité du prototype dans la réalité des opérations, de relever, s'il y a lieu, les lacunes des outils mis à leur disposition et de formuler des recommandations quant à leurs besoins. La première phase du projet PTT doit prendre fin en septembre 1990 et céder la place au deuxième volet.

L'avenir

Une fois l'essai opérationnel terminé, la version améliorée du PTT pourra être implantée dans certains services où des besoins précis auront été cernés. Pendant ce temps, la Direction de l'informatique évalue certaines composantes du PTT en vue de l'implantation éventuelle d'une trousse informatique plus modeste destinée à l'ensemble des traducteurs du Bureau. De plus, compte tenu de l'évolution rapide que connaît la micro-informatique, la Direction de l'informatique explore sans cesse de nouvelles avenues, à la recherche d'autres logiciels encore plus polyvalents. Elle travaille également à un ambitieux projet d'intégration du disque optique TERMIUM au PTT. Pour sa part, le CCRIT étudie plus particulièrement l'option réseau afin de donner au plus grand nombre possible de traducteurs accès à différentes bases de données électroniques communes. Les recherches vont bon train et déboucheront sur des résultats tangibles dans un avenir rapproché.

Annexe II

Traduction assistée par ordinateur (TAO)

Historique

Toujours en quête d'efficacité, le Bureau de la traduction étudie, depuis plusieurs années déjà, des systèmes de traduction automatique et, plus récemment, de traduction assistée par ordinateur. Le succès remporté par le projet TAUM MÉTÉO en 1977 et les grandes percées technologiques du début des années quatre-vingt ont amené le Bureau à faire l'essai de quelques logiciels de traduction assistée, dont ALPS, Microcat et Logos. Le choix du Bureau s'est toutefois arrêté sur le logiciel Logos en vue de son projet opérationnel échelonné sur cinq ans. Parmi les caractéristiques qui ont milité en faveur de Logos, mentionnons l'interaction au niveau du dictionnaire, la possibilité de modifier les composantes sémantiques en fonction du contexte et, finalement, la convivialité de l'outil dans son ensemble. Chaque année, le Bureau de la traduction doit traiter un volume énorme de textes techniques. Comme Logos se prête particulièrement bien à la traduction dans ce domaine, le Bureau sera en mesure de déterminer si le logiciel permet ou non de répondre à cette demande.

Essai en milieu opérationnel

Pour l'évaluation du logiciel en milieu opérationnel, des bancs d'essai, appelés sites TAO, ont été implantés dans six services de traduction. Il s'agit de la DSQ-Section informatique (Montréal), du QGDN (Ottawa), de la DSC-Section technique (Ottawa), des Transports (Ottawa), de l'Agriculture (Ottawa) et de la FMC (Saint-Hubert). Au total, quelque 35 traducteurs membres du personnel de soutien mettent chaque jour le logiciel Logos à l'épreuve dans des domaines d'application variés allant de l'informatique à la marine de guerre en passant par l'électronique. Outre Logos, les équipes disposent d'autres outils informatisés, dont LOTUS 1-2-3 pour la cueillette des données, WordPerfect pour la vérification orthographique et le traitement de texte, Kurzweil pour la saisie des textes sur support ordinolingue et les logiciels Keyword et Pride Local pour la conversion des fichiers de traitement de texte. Les équipes contribuent également à l'amélioration du logiciel Logos sous la forme de demandes ponctuelles présentées à la société Logos.

L'outil

L'équipement TAO de chaque site se compose d'une unité centrale Wang VS 5000, sur laquelle tourne le logiciel Logos, et des périphériques nécessaires. La majorité des traducteurs disposent d'un micro-ordinateur IBM compatible alors que les autres sont reliés au serveur grâce à des terminaux. Tous possèdent une imprimante au laser, et les deux traitements de texte privilégiés sont WordPerfect 5.0 dans le cas des micro-ordinateurs et Wang WP Plus dans celui des terminaux.

Le traducteur, appelé « taoïste » dans le jargon du métier, effectue son travail en quatre grandes étapes.

- **La pré-édition** : Le traducteur supprime les difficultés qui risquent de faire trébucher le logiciel.
- **La recherche terminologique** : Le système produit la liste des termes et locutions nominales qui ne figurent pas dans son dictionnaire informatisé. Quand le traducteur a trouvé les équivalents en faisant les recherches terminologiques et linguistiques nécessaires, il les intègre au dictionnaire ALEX en précisant le domaine à l'aide de codes et il rédige les règles sémantiques voulues grâce au programme SEMANTHA.
- **La traduction** : Il s'agit d'une opération purement automatique qui se déroule en arrière-plan. La sortie machine est versée dans un fichier de traitement de texte.
- **La post-édition** : Le traducteur révise à l'écran le produit brut avant de le livrer au client.

Gestion et coordination des projets de TAO

Des projets aussi ambitieux que ceux de la TAO exigent un solide encadrement. C'est ainsi que le comité directeur TAO, présidé par le sous-secrétaire d'État adjoint — Langues officielles et Traduction, dirige les projets de TAO, y compris ceux du PTT (poste de travail du traducteur) et du LATTER (l'atelier du terminologue). Pour planifier, coordonner et gérer les activités de TAO dites opérationnelles, le comité s'est doté d'un Secrétariat général TAO placé sous la responsabilité du directeur général — Planification, gestion et technologie (DGPGT). Le Secrétariat général regroupe des représentants de la Direction générale des opérations de traduction (DGOT), de la Direction de l'informatique et de la Direction de la technologie.

Pour mener à bien les recherches et les essais nécessaires dans les domaines linguistiques, informatique et technologique, on a créé le Centre d'expertise avec la collaboration des sociétés Wang et Logos.

L'avenir

La percée des logiciels de traduction automatique et de traduction assistée par ordinateur a bouleversé non seulement les méthodes de travail traditionnelles, mais la conception même de la profession. Dans sa vision de l'avenir, le Bureau de la traduction songe à l'intégration, en un seul et même système, de ses trois outils de haute technologie, à savoir le poste de travail du traducteur (PTT), la banque de terminologie TERMIUM et le logiciel de traduction Logos. Les progrès accomplis jusqu'à maintenant sont encourageants, et l'enthousiasme des deux principaux intervenants, le traducteur et le client, voue la TAO à un brillant avenir.

16

PTT-2 : la deuxième version du poste de travail du traducteur du CCRIT

Elliott MACKLOVITCH

Centre canadien de recherche sur l'informatisation du travail, Laval, Québec, Canada

Historique

Le poste de travail du traducteur[1], le PTT tel qu'on le désigne le plus souvent au Centre canadien de recherche sur l'informatisation du travail (CCRIT), est moins un produit qu'un *concept*. Le principe de base est le regroupement, sur la plate-forme relativement peu coûteuse qu'offre le PC, d'un ensemble de logiciels qui peuvent aider le traducteur à exécuter de façon plus efficace un certain nombre de tâches de routine périphériques[2].

La nature de ces tâches périphériques n'est pas la même pour tous les traducteurs. Nous avons entrepris le projet PTT en collaboration avec le Secrétariat d'État, et les traducteurs du Bureau de la traduction de ce ministère constituent notre clientèle cible. De 1988 à 1989, nous avons mis au point un prototype du PTT à l'intention des traducteurs du gouvernement fédéral, version installée sur un PC-AT qui comprenait,

1. Le terme « *traducteur* » désigne les personnes des deux sexes.
2. Nous entendons par tâches « *périphériques* » les taches qui s'écartent du rôle central du traducteur, lequel consiste par définition à reformuler le contenu du texte de la langue de départ en un texte correct et équivalent dans la langue d'arrivée. Dans ce sens, le PTT s'inscrit dans la catégorie des aides automatisées à la traduction humaine et non pas dans celle des véritables systèmes de traduction automatique, c'est-à-dire où la machine propose une traduction du texte de la langue de départ.

outre le logiciel intégrateur à fenêtres multiples (*DESQview*), les logiciels axés sur les fonctions et applications suivantes[3] :

- traitement de texte bilingue (*WordPerfect*),
- gestion personnelle ou locale de la terminologie (*Mercury/Termex*),
- repérage des mises à jour ou modifications de deux versions d'un fichier de traitement de texte (*CompareRite*),
- conjugaison de verbes en français (*Seconde Mémoire*),
- production d'une liste alphabétique des termes d'un texte et de leur fréquence et recherche de toutes les occurrences d'un terme donné avec affichage du contexte (*TextSearch*),
- conversion de fichiers d'un logiciel de traitement de texte à un autre (*SoftWare Bridge*).

Entre les mois de novembre 1989 et novembre 1990, ce prototype, appelé PTT-1, a fait l'objet d'un essai dans deux services du Bureau de la traduction, un à Ottawa et l'autre à Montréal. Treize personnes, traducteurs et réviseurs, ont participé à l'essai. Une petite équipe de spécialistes a procédé à une évaluation en bonne et due forme de l'essai, laquelle a surtout porté sur la convivialité et la fonctionnalité des diverses composantes du poste ainsi que sur les conséquences socio-organisationnelles de son implantation en contexte de travail au Bureau de la traduction[4]. L'équipe du CCRIT a tenu compte des résultats de cette évaluation lors de la conception du deuxième prototype du poste de travail du traducteur, le PTT-2.

Problèmes techniques

Plusieurs problèmes techniques sont associés au PTT-1 et certains étaient connus avant même que ne débute l'essai en milieu de travail. Comme nous l'avons déjà mentionné, la première version du poste de travail du traducteur, l'objet de l'essai, a pour base un PC-AT, équipé d'une carte d'extension mémoire et utilisant *DESQview* comme logiciel intégrateur, dans le but de permettre un accès simultané aux programmes d'application. Chacun des logiciels d'application s'exécute dans une fenêtre *DESQview* distincte dont la capacité de mémoire ne peut dépasser celle de la zone de mémoire conventionnelle contiguë la plus large encore disponible une fois chargés les systèmes *DOS* et *DESQview* ainsi que les modules de commande de la souris et de l'écran pleine page. En conséquence, un espace mémoire insuffisant (soit environ 409 ko dans le cas du PTT-1 sur PC-AT) restreint plusieurs des possibilités du poste de travail. Ainsi, la capacité limitée de la mémoire du PTT-1 ne permet pas de charger dans la même fenêtre le logiciel de gestion de glossaires du poste de travail, *Mercury/Termex*, et

3. Pour plus de détails sur cette première version du PTT, voir E. Macklovitch, « An Off-the-Shelf WorkStation for Translators », dans *Proceedings of the 30th Annual Conference of the ATA*, D. Hammond, (editor), Washington, D.C., 1989, p. 491-498. (Également publié en français sous forme du rapport technique du CCRIT.)
4. Voir R. Lavoie et F. Lapointe, *Poste de travail pour la traduction : Évaluation de la convivialité*, rapport technique du CCRIT, janvier 1991.

WordPerfect version 5.0. En effet, le logiciel *Mercury/Termex* a été conçu en tant que programme résidant en mémoire et lorsqu'il n'est pas exécuté dans la même fenêtre que le traitement de texte, on perd l'intéressante fonction couper-coller. Le manque de mémoire empêche aussi le chargement de *Findit*, le logiciel d'interrogation donnant accès à la version sur CD-ROM de *Termium*, la banque de terminologie du Secrétariat d'État.

Ces deux types de problèmes peuvent être résolus par l'adoption d'un PC de type 386 comme plate-forme matérielle et par l'utilisation d'un programme de gestion de mémoire d'expansion comme le *QEMM-386* de *Quarterdeck*. Le programme *QEMM* augmente la capacité de la plus grande zone de mémoire conventionnelle disponible en remplissant les espaces adresses inutilisés de la mémoire haute, c'est-à-dire les adresses situées entre 640 Ko et 1024 Ko[5]. Le chargement de plusieurs des programmes de commande et petits programmes utilitaires du PTT dans la mémoire haute permet de libérer suffisamment d'espace dans la mémoire conventionnelle pour que *Findit* puisse être exécuté dans une fenêtre *DESQview*; cette plus grande capacité de mémoire permet également d'exécuter *Termex* et *WordPerfect* dans la même fenêtre et du même coup de récupérer la fonction couper-coller.

Un autre problème technique révélé par l'essai a trait à l'écran du poste de travail. Dans les consultations qui ont précédé la définition des spécifications techniques du PTT-1, les traducteurs avaient exprimé le souhait de travailler avec un écran pleine page permettant l'affichage de plus de 25 lignes de texte à la fois et de plus d'une fenêtre. En réponse à cette demande, l'équipe du CCRIT a proposé *The Genius*, écran monochrome à haute résolution fabriqué par *Micro Display Systems*. Dès le départ, les participants à l'essai ont formulé des commentaires très négatifs sur l'écran *Genius*, car malgré l'avantage que présente sa plus grande taille, la grande majorité des utilisateurs jugeaient les caractères beaucoup trop petits en mode 66 lignes. Les participants ont continué de se montrer insatisfaits de l'écran *Genius* même après la reconfiguration du PTT en mode 41 lignes. En effet, les autres problèmes soulevés ne pouvaient être corrigés aussi facilement, notamment le scintillement constant sur presque tous les modèles et le mauvais alignement du cadre d'image sur certains. De plus, l'écran *Genius* n'est pas compatible VGA, ce qui signifie qu'il n'est pas possible de tirer avantage de la capacité du *DOS 3.3* d'afficher les majuscules accentuées. Compte tenu de tous ces problèmes, nous avons décidé de remplacer l'écran *Genius* par un écran couleur VGA 16 pouces, mais nous sommes toujours à la recherche d'un écran monochrome pleine page dont la fiabilité soit reconnue afin de pouvoir offrir une solution de rechange.

Nouvelles fonctions

Une des principales différences entre la première et la deuxième version du poste de travail du traducteur est que le PTT-2 fonctionne en réseau, tandis que le PTT-1 est

5. Il existe un programme *Quarterdeck* similaire pour les PC-286, mais il ne convient pas à tous les ordinateurs de ce type, dont ceux en place au Bureau de la traduction.

installé sur un PC autonome. Un certain nombre de raisons ont motivé un changement d'orientation aussi important, plusieurs venant des préoccupations exprimées par les participants à l'essai lors des entrevues d'évaluation. Si nous reformulons les remarques des traducteurs à un niveau légèrement plus abstrait, nous pouvons dire qu'un réseau favorise généralement un partage plus rationnel et plus économique des ressources, comparativement à un poste autonome. Parmi ces ressources, on compte les principaux fichiers de données ou les programmes importants ainsi que certains appareils périphériques. Un réseau permet aussi aux utilisateurs de bénéficier de services communs comme une copie de sauvegarde périodique ou l'installation de nouvelles versions de logiciels; ces tâches deviennent moins répétitives lorsqu'elles sont exécutées par l'intermédiaire d'un serveur central.

Si nous en venons aux points précis favorables à un réseau qui ont été exprimés par les participants à l'essai, plusieurs traducteurs sont d'avis qu'un réseau local faciliterait l'archivage, la recherche et l'extraction de versions électroniques de textes déjà traduits, non seulement les leurs, mais ceux de l'ensemble de la section de travail. Comme nous le verrons plus loin, il s'agit d'une amélioration des plus souhaitables par rapport au système actuel qui consiste à classer les copies papier dans des classeurs volumineux. D'autres ont fait remarquer qu'un réseau rendrait également plus facile la tâche de mise à jour du fichier électronique de terminologie de la section. À l'heure actuelle, il s'agit d'un travail relativement ardu qui consiste à extraire tous les nouveaux enregistrements (fiches terminologiques) ajoutés au glossaire personnel de chaque utilisateur depuis la dernière mise à jour; à fusionner ces enregistrements (après vérification) dans le glossaire commun de section; à convertir ce glossaire en un dictionnaire d'arrière-plan non modifiable; et enfin, à recharger ce nouveau fichier dans la machine de chaque traducteur. Une installation en réseau peut simplifier au moins la première et la dernière de ces étapes, c'est-à-dire la collecte des fiches individuelles et la redistribution de la version à jour du glossaire de section. Dans le même ordre d'idées, le réseau peut faciliter l'organisation d'autres aspects du travail qui touchent l'ensemble de la section, comme la répartition en tranches des textes à traduire et la diffusion de messages électroniques.

Outre les fichiers de données et les programmes communs, le réseau de PTT du CCRIT offre aussi un accès partagé à un certain nombre d'appareils périphériques, dont une imprimante au laser de haute qualité et un lecteur de CD-ROM (Compact Disk-Read Only Memory). Ce lecteur permet d'interroger simultanément de plusieurs postes la version sur CD-ROM de *Termium*, l'énorme banque de terminologie du Secrétariat d'État, ainsi que d'autres ressources précieuses[6] également sous forme de disques compacts, dont le nombre ne cesse d'augmenter. Notre réseau est aussi doté d'une unité de sauvegarde sur bande qui réduit les risques de perte accidentelle de fichiers électroniques importants en permettant aux traducteurs d'archiver ces fichiers dans des répertoires du serveur, lesquels font automatiquement l'objet de copies de

6. *Le Grand Robert* par exemple, probablement un des dictionnaires de la langue française les plus complets. De plus, les versions sur disque compact de ces ressources peuvent faire l'objet de recherches complexes, impossibles dans les versions publiées, c'est-à-dire non limités à la recherche alphabétique par vedette.

sécurité à intervalles périodiques. Dans les sections qui suivent, nous analysons un peu plus en détail deux des principaux changements mis en œuvre dans notre réseau de PTT.

Termium pour l'ensemble du réseau

Au moment de la conception du projet de poste de travail du traducteur, nous voulions offrir l'accès à la version sur CD-ROM de *Termium* à partir d'une fenêtre de *DESQview*, afin que le traducteur puisse, à l'aide d'une simple touche, passer de son logiciel de traitement de texte à la banque de terminologie et entrer directement dans son texte l'information ainsi obtenue. Cependant, comme nous l'avons déjà souligné, les limites de mémoire du PTT-1 sur PC de type 286 ne permettaient pas une telle installation. Le problème est résolu en grande partie par l'utilisation d'un PC de type 386. Avec une mémoire conventionnelle accrue, le logiciel d'interrogation peut être exécuté sans problème dans une fenêtre *DESQview*, à condition bien sûr que chaque poste soit relié à un lecteur de CD-ROM. L'installation en réseau de tous les postes de travail de la section permet un accès partagé à un seul lecteur de CD-ROM, une utilisation sans doute plus rationnelle des ressources. Toutefois, il faut à cette fin ajouter un autre logiciel (en plus de *Microsoft CD-ROM Extensions*) à chaque poste du réseau, afin de convertir le lecteur de CD-ROM relié à un appareil en un périphérique à accès partagé. Le progiciel que nous avons choisi pour notre réseau s'appelle *OPTI-NET*[7]. Il en existe deux versions, selon le nombre de postes qui consultent le disque compact; il est compatible avec divers types de réseaux de PC; et il a pour avantage de pouvoir être installé dans un serveur de CD-ROM non spécialisé, c'est-à-dire un ordinateur dans lequel d'autres applications peuvent être exécutées en concurrence. Bien que l'accès au disque compact puisse être considéré comme relativement lent, selon les résultats des premiers tests que nous avons effectués et qui ont porté sur l'interrogation simultanée de *Termium* depuis trois postes de travail, les temps de réponse restent acceptables.

Gestion de documents

Une des premières choses que fait un traducteur du Bureau lorsqu'il reçoit un texte est de vérifier dans les classeurs de la section si le texte n'a pas déjà été traduit ou s'il existe un texte de même nature dont la terminologie pourrait être utile au nouveau texte. Chaque section du Bureau est responsable de l'organisation de son système de classement, et certaines peuvent être plus méthodiques que d'autres. Supposons, à titre d'exemple, que les traductions sont classées par ministère client dans toutes les sections. Il serait donc tout à fait possible que le traducteur ne découvre pas un texte d'une très grande pertinence parce qu'il a été traduit pour un autre ministère (à moins que la

7. *OPTINET* n'est pas la seule possibilité; pour de plus amples renseignements sur la question de l'exploitation en réseau des disques compacts, voir *PC Magazine*, vol. 9, n° 4, 27 février 1990.

section ne se soit donné la peine d'établir un système élaboré de correspondance). En tant qu'ancien traducteur du Bureau, je suis porté à croire que plus souvent qu'autrement, les traducteurs ne réussissent pas à retrouver les documents très pertinents qui dorment dans la mémoire collective de la section et n'ont d'autre choix que de retraduire une part importante d'un travail déjà fait.

Lorsque nous avons conçu la deuxième version du poste de travail du traducteur, un de nos objectifs était de démontrer que ce genre de problème pouvait être résolu en partie grâce à une base de données relationnelles standard. Au lieu de définir nous-mêmes la structure d'une telle base de données, nous avons décidé de nous fonder sur la formule « *Demande de traduction* », un bordereau d'envoi qui accompagne chaque texte destiné au Bureau de la traduction. Sur cette formule, on trouve des renseignements comme le titre et le nom de l'auteur du document à traduire, le nom du ministère et de la division, la nature du document (par ex. rapport, compte rendu, description de fonctions), la date de réception de la demande, le nombre de mots et un numéro d'identification unique. Une fois sa traduction achevée, un traducteur travaillant sur PTT-2 rappelle la formule (version électronique) à l'aide d'une macro exécutée dans son programme de traitement de texte, la remplit et peut même ajouter un nombre illimité de mots clés qui décrivent bien le contenu du texte. La sauvegarde de la formule à l'aide de la même macro entraîne son classement automatique dans le répertoire approprié du serveur du réseau en vue de la mise à jour ultérieure de la base de données[8]. Lorsqu'un traducteur veut vérifier si le texte qu'il vient de recevoir a déjà été traduit, il peut interroger la base de données à partir de son poste, en soumettant toute combinaison voulue de clés de recherche tirées du texte de départ. Il peut aussi faire afficher une liste des éléments contenus dans chacun des index de la base et, au moyen d'un système de menus, composer des demandes d'information structurées du genre : « *Avons-nous des textes portant sur le chômage qui auraient été soumis par le client X et traduits par le traducteur Y entre les mois de septembre et de décembre de l'année* Z ». Si la base de données trouve des enregistrements qui concordent avec les renseignements demandés, elle affiche les formules de demande de traduction correspondantes; et si, après examen de ces renseignements, le traducteur juge qu'une de ces demandes s'applique au travail en cours, il note simplement le numéro et s'en sert pour rappeler le texte intégral et sa traduction dans le répertoire des textes déjà traduits qui se trouve dans le serveur.

Rappelons que le module de démonstration qui vient d'être décrit fait partie d'un prototype non encore utilisé par des traducteurs en contexte opérationnel. La base de données que nous avons constituée pour les besoins de la démonstration est relativement petite, mais rien ne nous porte à croire que son efficacité serait moindre si elle était élargie de manière significative. Nous avons choisi une base de données structurée de préférence à un logiciel de recherche plein texte parce que nous pensons que plus la

8. Nous utilisons les macros du traitement de texte pour l'ajout de nouveaux enregistrements dans la base de données pour simplifier la vie du traducteur et lui éviter d'avoir à apprendre un autre éditeur de texte comme celui fourni avec la base de données. La mise à jour proprement dite de la base de données devrait être confiée à un responsable désigné, qui l'effectuerait à partir du programme de gestion de la base de données.

taille de la base de données s'élargit, plus le nombre de réponses non pertinentes risque d'augmenter. Quoi qu'il en soit, le module n'a d'autre but que de démontrer ce qui peut être fait, il ne constitue donc qu'une solution possible au problème de la gestion des documents, une fois que tous les textes sont archivés sous forme électronique dans un réseau local[9].

Réduction de l'hétérogénéité

Comme nous le précisions au début de ce document, notre projet de poste de travail repose avant tout sur le regroupement d'un ensemble d'applications logicielles, chacune visant à aider le traducteur à exécuter certaines tâches périphériques, de routine. On pourrait envisager le développement à partir de zéro d'un tel ensemble de programmes, ce qui aurait pour avantage de permettre un degré élevé de cohérence et d'uniformité entre les fonctions similaires des diverses applications. Nous avons plutôt choisi de repérer des progiciels commerciaux existants qui, même s'ils n'ont pas été conçus pour les tâches de traduction, pourraient néanmoins être utiles à la traduction humaine. L'avantage évident d'une telle façon de procéder est qu'elle évite les coûts élevés de la programmation et de l'essai intégral de chaque élément; en outre, elle permet d'assembler et de livrer un prototype opérationnel en beaucoup moins de temps. Bien sûr, le défi consiste à intégrer ces programmes disparates, avec les commandes et le mode de fonctionnement qui leur sont propres. La tâche aurait été moins ardue si nous avions adopté comme plate-forme un appareil comme le Macintosh dont le système d'exploitation oblige les concepteurs de logiciels à fournir des produits d'une assez grande uniformité. Cependant, notre principal client avait dès le départ imposé la contrainte du choix de la norme PC, surtout par souci de réduire autant que possible les incompatibilités entre les systèmes de traitement de texte des ministères auxquels il offre ses services. Par contre, d'autres facteurs militent en faveur du PC, comme un éventail plus large de logiciels parmi lesquels choisir, et le fait que *Termium III*, la banque de terminologie du gouvernement et un des éléments essentiels de notre poste de travail, ne fonctionne qu'avec le *DOS*.

Comme nous l'avons vu, *DESQview*, le logiciel intégrateur commercial utilisé pour la version PTT-1, atténue dans une certaine mesure le problème de l'hétérogénéité des programmes, en offrant ce que l'on a appelé l'*intégration d'accès*. En d'autres termes, l'utilisateur peut charger l'application qu'il a choisie simplement en la sélectionnant dans le menu offert; il peut ensuite se déplacer ou transférer du texte d'une fenêtre d'application à une autre, tout aussi simplement. Toutefois, à part cette facilité d'accès, *DESQview* n'a pas grand-chose à offrir : une fois la fenêtre d'application ouverte, l'utilisateur doit connaître le fonctionnement de chaque programme et se rappeler les commandes du clavier ou du menu à utiliser dans chaque cas.

9. Même pour notre client particulier, un module distinct de gestion de documents n'est pas nécessairement la meilleure solution. L'idéal serait son intégration au Système d'information de gestion (SIO) du Bureau de la traduction qui contient déjà sensiblement les mêmes informations, tirées de la Demande de traduction. Inutile de dire qu'il s'agit là d'une entreprise de taille qui déborde largement le cadre de notre présent mandat.

De ce fait, le fardeau sur le plan de l'apprentissage est relativement lourd à porter pour l'utilisateur et un dilemme important se pose pour le concepteur du poste de travail. Plus nous ajoutons d'applications au poste de travail — dans le but d'aider le traducteur à exécuter un plus grand nombre de tâches de façon plus efficace —, plus il devient difficile pour le traducteur de retenir les diverses commandes des logiciels et d'exploiter toutes leurs possibilités. Cette tension a été observée de façon très claire dans l'évaluation de l'essai en milieu de travail du PTT-1. En effet, les évaluateurs ont constaté que les participants connaissaient à peine certains des programmes de leur poste de travail, malgré le fait qu'ils avaient reçu une formation de base sur ces applications au début de l'essai. En revanche, les applications que les traducteurs maîtrisaient le mieux à la fin de l'essai étaient celles-là mêmes qu'ils estimaient avoir à utiliser le plus dans leur travail de tous les jours — par exemple *Mercury/Termex* pour la consultation de leurs glossaires sous forme électronique, et *5M-Verbes*, le programme de conjugaison de verbes français. Soulignons que si les participants n'ont pas réussi à tirer profit des avantages des autres programmes, ce n'est certes pas par paresse ou manque de bonne volonté. Soit ils étaient d'avis que les possibilités offertes par ces programmes ne convenaient pas vraiment à leurs tâches et donc ne justifiaient pas l'effort d'apprentissage, soit le poste de travail exigeait déjà trop d'efforts de leur part. Pour être juste à l'égard des participants, il importe de préciser que plusieurs ne s'étaient jamais servis d'un ordinateur avant qu'on ne décide qu'ils participeraient à l'essai. Quant aux autres, la seule formation reçue avant que l'essai ne débute se limitait à quelques brèves séances de cours sur l'utilisation du traitement de texte. À cet égard, il convient de citer une autre conclusion intéressante des évaluateurs, à savoir que le degré d'appréciation des diverses composantes du poste de travail était directement relié au degré de familiarisation avec *WordPerfect*. On peut donc penser que ceux qui venaient seulement d'apprendre les rudiments de ce puissant logiciel de traitement de texte étaient trop préoccupés par cet apprentissage pour vraiment être en mesure d'explorer toute autre nouvelle application.

À court terme

Les résultats de l'évaluation démontrent de façon claire que les besoins en formation avant-essai des participants avaient nettement été sous-estimés. Aussi, comme solution au problème, le Bureau de la traduction a-t-il décidé d'organiser des séances de formation supplémentaires à l'intention des participants à l'essai appelés à continuer à travailler sur une version améliorée du PTT-1[10]. De plus, les applications que les participants n'avaient pas jugé essentielles à leurs tâches, en particulier celles au langage de commande abstrus comme *TextSearch*, ont été éliminées de cette version améliorée ainsi que du PTT-2 mis au point par le CCRIT. D'autres programmes, reconnus comme étant d'une utilité moyenne mais qui sont rarement consultés, comme *Software Bridge*,

10. Cette version améliorée, appelée Pl.I, est également installée sur un PC de type 386, mais contrairement au nouveau PTT-2, il ne fonctionne pas en réseau et ne comprend pas un module de gestion de documents.

sont retenus, mais ils seront installés sur un seul appareil par section. Enfin, soulignons que toute nouvelle application possible — même celles pour lesquelles les traducteurs ont exprimé un vif intérêt, comme un correcteur orthographique capable de vérifier l'application des règles d'accord en français[11] — devra être analysée sous toutes ces coutures avant d'être adoptée, et la préférence sera accordée aux produits dont l'interface est conviviale. Le même raisonnement s'applique, et à plus forte raison, aux autres programmes utilitaires qui ne présentent qu'un intérêt secondaire pour un traducteur professionnel. En résumé, bien que nous introduisions des changements en mettant dans le PTT-2 de nouvelles versions des logiciels qui étaient dans le PTT-1, nous voulons avant tout limiter le nombre de ces applications, du moins aussi longtemps que les utilisateurs ne maîtriseront pas à fond le fonctionnement des éléments centraux.

Soulignons que cette réduction du poste de travail du traducteur ne doit pas être interprétée comme la conséquence d'une mauvaise conception de base du prototype original; au contraire. Cette mesure est tout à fait conforme à notre façon de voir le poste de travail comme un ensemble évolutif de logiciels; il est on ne peut plus souhaitable que les utilisateurs finals soumettent leurs commentaires aux concepteurs afin qu'ils puissent servir à modifier et à améliorer la configuration originale. Dans ce projet conjoint, le CCRIT propose des applications jugées utiles à la traduction, mais il appartient aux traducteurs eux-mêmes d'évaluer la pertinence et la facilité d'utilisation de ces aides.

À moyen terme

De toute évidence, les mesures décrites dans la section précédente ressemblent fort à un traitement palliatif. À long terme, si l'on veut résoudre de façon définitive le problème de l'hétérogénéité des applications qui composent le poste de travail, il faudra faire plus que simplement réduire le nombre des composantes. La solution viendra peut-être d'un produit comme *Windows* version 3.0 de *Microsoft*, l'interface graphique sur PC qui est en train d'envahir le marché. En plus de posséder tous les avantages des interfaces graphiques multifenêtres — notamment la représentation vraiment fidèle de tous les attributs de mise en page et une utilisation intégrale de la souris et des icônes — *Windows* fait ce que fait *Finder* pour le *Macintosh* : il oblige les concepteurs de logiciels à respecter un ensemble de conventions de base en matière de création de menus au moment de définir leur répertoire de commandes. Le fait ne peut que contribuer à diminuer nettement le fardeau des utilisateurs en ce qui concerne l'apprentissage et la maîtrise de nouvelles applications.

---

11. Par exemple, *Hugo Plus* et *Ortograf+*. Si ces logiciels donnaient des résultats fiables du point de vue de la détection des erreurs grammaticales, il vaudrait la peine de les intégrer au poste de travail. Cependant, d'après les premiers résultats des tests auxquels j'ai soumis ces produits, il semble que ce ne soit pas le cas. Toute évaluation poussée doit comporter une étude du nombre de fausses erreurs grammaticales signalées par ces programmes ainsi que du type et du nombre de données grammaticales que les utilisateurs doivent introduire pour élargir le contenu de leurs dictionnaires.

À ce jour, très peu des applications qui composent notre poste de travail sont offertes en version *Windows*, mais la société qui fabrique *WordPerfect*, l'élément de traitement de texte central qui est de fait devenu la norme dans l'administration fédérale, a annoncé son intention de sortir sa version *Windows* pour le premier trimestre de 1991. Il semble qu'une pression de plus en plus forte s'exerce sur l'ensemble des concepteurs de logiciels — même dans le cas des applications « *linguistiques* » de moindre envergure du genre de celles que nous utilisons dans notre poste de travail — afin qu'ils produisent une version *Windows*, ne serait-ce que pour conserver leur part du marché. En attendant, nous avons commencé à monter une version *Windows* du poste de travail, surtout à des fins de démonstration et d'évaluation. Le traitement de texte choisi est *Word for Windows*, mais toutes les autres composantes du PTT-2 fonctionnent avec le *DOS*. Pour le moment, nous avons l'impression qu'il n'y aurait pas grand avantage à exécuter ces applications *DOS* dans le cadre de *Windows*; comparativement à l'environnement *DESQview*, leur chargement est plus lent et leur rendement général inférieur dans l'ensemble. De plus, comparativement à *DESQview*, *Windows* exige d'avantage de mémoire et un processeur plus rapide; il est, en outre, beaucoup plus difficile à configurer de façon optimale.

Pour être juste envers *Windows*, il faudrait l'évaluer avec des applications spécialement conçues pour cette interface. Et bien qu'il n'entre pas dans notre mandat de comparer et d'évaluer des systèmes de traitement de texte, nous ne faisons de tort à personne en mentionnant que *Word for Windows* est un outil extrêmement puissant et agréable à utiliser, qui comporte nombre des plus intéressantes caractéristiques de mise en page normalement associées aux logiciels d'édition assistée par ordinateur. Cette impression positive est renforcée par une autre application commandée spécialement par le CCRIT pour le projet de poste de travail : un module de dictionnaires électroniques regroupés sous le nom de *InContext*. Mis au point par *ESP Inc.*, une entreprise de Toronto, *InContext* ne fonctionne qu'en environnement *Windows*, comme programme autonome ou comme option du menu principal de *Word*. Par l'entremise d'une simple fenêtre, *InContext* offre actuellement l'accès à trois bases de données lexicales : le *Gage Canadian English Dictionary*, le *Multidictionnaire des difficultés de la langue française* et le *Hayakawa Thesaurus of English*. Bien que le produit soit encore à l'étape de l'essai pilote, on prévoit déjà y ajouter d'autres ressources lexicales, dont un dictionnaire bilingue anglais-français. Lorsque *InContext* est chargé dans la même fenêtre que le traitement de texte, l'utilisateur n'a qu'à sélectionner dans son texte le terme qu'il désire chercher, choisir l'option *InContext* figurant dans le menu et indiquer la base de données qu'il veut consulter; quelques secondes plus tard, la définition complète du terme s'affiche à l'écran — presque identique à celle qui figure dans la version publiée de l'ouvrage de référence[12]. Si l'utilisateur désire copier des parties de la définition dans son texte, il n'a qu'à les sélectionner à l'aide de la souris, à choisir l'option « *cut* » du menu d'édition et ensuite l'option « *paste* » du menu

12. Chose surprenante, *ESP* a eu de la difficulté à reproduire les divers types de polices de caractères qui permettent notamment de distinguer la définition elle-même des exemples, abréviations, etc. On aurait pu penser que *Windows* serait un environnement approprié à la reproduction de ces caractéristiques particulières, mais il semble que les problèmes ne sont pas tous résolus.

Word — comme pour toutes les autres applications *Windows*. La simplicité de l'interface est ce qui rend le produit si intéressant; sans compter également le fait qu'il y a une interface unique pour plusieurs dictionnaires électroniques distincts[13]. Les traducteurs qui ont vu *InContext* à l'œuvre se sont montrés très impressionnés, tout en exprimant le souhait qu'on y ajoute d'autres dictionnaires.

Jusqu'ici, *Word* et *InContext* sont les seules applications *Windows* qui ont été installées sur cette nouvelle version du poste de travail du traducteur; d'autres viendront s'y ajouter à mesure qu'elles seront disponibles. Les résultats déjà obtenus sont cependant très encourageants. Avec une interface comme *Windows*, le poste de travail peut offrir plus que la simple intégration de l'accès à ses composantes, mais aussi ce que l'on a appelé l'intégration complète des applications. Nous sommes portés à croire qu'il s'agit d'une solution viable au dilemme décrit au début de ce chapitre, traçant la voie au futur PTT-3.

Conclusion

Comme nous l'avons souligné au début de ce document, le PTT n'est pas un produit fixe, mais bien un ensemble de logiciels dont la composition peut varier en fonction des besoins particuliers des traducteurs qui l'utilisent. Le concept du PTT est suffisamment souple pour accepter diverses réalisations, de sorte que les fonctions que nous venons de décrire pourraient être exécutées tout aussi bien avec des logiciels différents ou même avec un autre type de plate-forme matérielle[14]. Ceci étant dit, la liste qui suit présente un résumé des composantes matérielles et logicielles de notre PTT-2.

Matériel

1) Poste de travail
   - PC-386 (*Zénith Z-386116*), doté de 4 Mo de mémoire vive,
   - écran VGA 16″ couleur (*NEC Multisync 4D*),
   - souris (*Logitech serial*),
   - unité interne de sauvegarde sur bande (*Wangtek PC-36*).
2) Serveur du réseau
   - PC-386 (*Zénith Z-386/16*) doté d'un disque dur de 150 Mo et de 4 Mo de mémoire vive,
   - plaque tournante active à 4 portes (*Pure Data*),
   - cartes d'interface *ARCnet* 8 bits (*Pure Data*).
3) Serveur de CD-ROM
   - PC-386 (*Ogivar SX386*),
   - lecteur externe de CD-ROM (*Hitachi 15035*).
4) Imprimante au laser (*HP LaserjetIII*).

13. En comparaison, chaque base de données lexicales archivée sur un CD-ROM possède son propre logiciel d'extraction et sa propre interface.
14. Pour une description d'un PTT sur *Macintosh*, voir « La solution Thibodeau » dans *Circuit*, n° 27, septembre 1989, p. 33-34.

Logiciels

- *Novell Advanced Netware 286 v2.15*,
- *DOS* version 3.3,
- *DESQview* 386 (y compris *QEMM-386*),
- *WordPerfect 5.1*,
- *Mercury/Termex 2.0* (y compris le dictionnaire anglais-français *Harraps*),
- *CompareRite 3.15*,
- *Verbatex* (pour la conjugaison des verbes français),
- *OPTI-NET* version 1.20 (pour accéder à une CD-ROM en réseau local),
- *Findit 5.0* (logiciel d'interrogation de *Termium* sur CD-ROM),
- *SY-TOS* (logiciel de l'unité de sauvegarde sur bande).

17

Trois scénarios possibles pour la station de travail du traducteur en l'an 2001

Poul Søren KJÆRSGAARD

Université d'Odense, Odense, Danemark

Un humoriste de mon pays disait il y a quelques décennies que rien n'est plus difficile que de prévoir l'avenir.

Puisque je ne me crois pas doué pour ce genre d'activité, mon intervention se propose d'esquisser des scénarios qui s'enracinent dans le présent ou même dans le passé récent. Je trouve légitime de faire cela, car il me semble que la décennie que nous venons d'entamer ne verra pas d'évolution spectaculaire dans le domaine de la traduction automatisée[1], mais au contraire une élaboration des idées qui sont actuellement en cours. Autrement dit, il me semble que la situation actuelle se caractérise très bien par ce qu'écrivait le philosophe allemand Leibniz : *Le présent est gros de l'avenir*[2].

Récurrence historique

L'année dernière, on a vu, au Danemark, la commercialisation d'un logiciel de traduction assistée par ordinateur qui s'appelle *Winger92*.

1. J'entends par ce terme aussi bien la traduction dite automatique que la traduction assistée par ordinateur, qu'elle soit dirigée par l'homme (THAM) ou par l'ordinateur (TAAH).
2. Cette idée, qu'on trouve dans les *Nouveaux essais sur l'entendement humain*, fut reprise par Voltaire dans son *Dictionnaire philosophique*.

C'est un événement caractéristique à plusieurs niveaux et je crois en outre que le cas est suffisamment typique ou généralisé pour qu'il permette d'en tirer des leçons.

- Il s'agit d'une toute petite entreprise (moins de 10 employés).
- Aucun linguiste n'y a travaillé à plein temps.
- Par conséquent, une idée directrice de ce logiciel est d'attaquer le processus de traduction par le biais de la technologie. Le principal atout du produit réside en effet dans une nouvelle technique pour faire un dictionnaire bilingue très compact, de façon qu'il n'absorbe pas toute la mémoire vive de l'ordinateur. Ainsi, le logiciel fonctionne sur PC. Outre un dictionnaire bilingue, le logiciel ne contient aucun savoir linguistique, et on en fait même une vertu.

Une autre idée directrice du logiciel est de traduire mot à mot ou phrase par phrase (par phrase, on entend ici un groupe de mots).

Une session typique de ce logiciel se déroule de la façon suivante : le texte à traduire (le texte source) est lu par l'ordinateur, puis est traduit mot à mot.

Chaque fois qu'un mot du texte source équivaut — selon le dictionnaire bilingue — à plusieurs mots ou expressions dans la langue d'arrivée, le programme s'arrête en principe pour demander conseil au traducteur. En principe seulement, car les concepteurs se sont vite rendu compte que de nombreux mots sont ambigus. Mentionnons quelques exemples anglais et français :

MAN	verbe et nom
GREEN	adjectif et nom
CAN	auxiliaire et nom
MONTRE	verbe (présent) et nom polysémique
FERME	verbe (présent), adjectif et nom
LE	article et pronom personnel

Pour accélérer le rythme du programme, donc pour remédier à ce problème (que d'autres considéreraient comme une richesse des langues dites naturelles), les concepteurs ont introduit le principe du *longest string matching* : le programme va établir l'entrée la plus longue du dictionnaire qui contient le mot en question. Exemples :

NIGHT	SOIR/SOIRÉE NUIT
ADOPTER	VEDTAGE (en lov [un projet de loi]) ADOPTERE (et barn [un enfant])
CAN	PEUT (auxiliaire) BOÎTE DE CONSERVES

Dans ces exemples, un contexte plus large pourrait être trouvé en juxtaposant, selon les cas, une préposition ou bien un article :

AT NIGHT	(la nuit)
LAST NIGHT	(hier soir)
ALL NIGHT	(toute la nuit)
A CAN	(une boîte)
THE CAN	(la boîte)

La traduction mot à mot implique qu'il s'agit d'un logiciel de première génération (ou même avant).

Mais là où, progressivement, les logiciels de première génération introduisaient des règles locales, basées essentiellement sur l'ordre des mots, donc sur la compatibilité structurale ou syntaxique entre un mot et son prédécesseur ou bien son successeur, les cas typiques étant *article-nom* ou *pronom personnel-verbe fini*, il n'y a même pas de règles (ni indication dans le dictionnaire de la partie du discours à laquelle appartient le mot en question), on s'appuie exclusivement sur la technologie en incluant toutes les formes fléchies de tel nom ou de tel verbe dans le dictionnaire ainsi que l'entrée de tous les mots composés. Grâce à une méthode de stockage très comprimé, le dictionnaire ne dépasse pas les limites d'un PC ordinaire, et grâce à une technique de recherche très rapide, la vitesse de croisière de ce logiciel atteint quelque soixante mots traduits par minute.

Pour pallier les problèmes dus au fait qu'une chaîne de caractères (un mot) appartient à deux catégories, dont une très fréquente et l'autre relativement rare, les concepteurs s'imaginent que les problèmes peuvent être résolus à l'aide de mots juxtaposés. Mais on trouve sans difficulté des exemples dans lesquels ce ne serait pas le cas :

LA DIRECTION EST RESPONSABLE
LA DIRECTION EST DU VENT

Puisque la base de données permet un stockage très dense, les concepteurs proposent qu'on entre toutes les phrases dans lesquelles on trouve NIGHT. Pour l'occasion, on a inventé un principe qu'on dénomme, avec un abus de langage, ATA, acronyme pour *Acquired Translation Ability*. Cela veut dire, en toute simplicité, que le logiciel est assorti d'un module assez souple pour faciliter l'entrée et le stockage de nouveaux mots/phrases dans la base de données.

Mais les concepteurs n'ont certainement pensé ni au facteur temps ni au caractère fastidieux de ce travail dans le cas des langues flexionnelles commes les langues romanes (pour ne pas parler du finnois) ou dans le cas des langues germaniques qui combinent aisément les mots :

LEBENSVERSICHERUNGSGESELLSCHAFTSANGESTELLTER
(un employé d'une compagnie d'assurance-vie)
(exemple repris de Martin Kay)

Mais l'approche contient également des difficultés d'ordre théorique. Même si l'on avait stocké les phrases

ADOPTER LE BÉBÉ
ADOPTER LE PROJET DE LOI

pour désambiguïser le mot ADOPTER (en l'occurrence, par rapport aux langues allemande et danoise), cette chaîne serait décomposée en cas de passivation de la phrase :

LE PROJET DE LOI FUT ADOPTÉ
LE BÉBÉ FUT ADOPTÉ

de façon que le principe de juxtaposition ne soit plus utilisable.

Enfin, si les mots, pris individuellement, sont ambigus, les phrases le sont parfois aussi, comme dans l'expression anglaise :

AFTER ALL	malgré tout, tout compte fait
AFTER ALL THAT TROUBLE	après toutes ces difficultés

On aura compris — je l'espère — qu'il s'agit d'un logiciel très primitif. Si j'ai choisi de porter à l'attention d'un public étranger l'existence de ce logiciel (sans grand intérêt, ni sur le plan linguistique ni sur le plan technique), ce n'est pas pour discourir sur les *primitifs danois* en matière de linguistique informatique. Pour autant que je sache, il existe des logiciels plus ou moins similaires dans les pays voisins, et je crois (je crains) que les années à venir verront la commercialisation d'autres logiciels de ce type. La plupart de ces logiciels connaîtront sans doute une existence éphémère.

La raison est plutôt que la commercialisation de ce logiciel met en lumière un double divorce, d'un côté entre le potentiel linguistique en TA (qui, certes, est loin d'être parfait, mais néanmoins infiniment supérieur à Winger92) et une certaine gamme de logiciels commercialisés, dont Winger92. Et de l'autre côté, un divorce entre ce qui nous préoccupe, nous linguistes et traducteurs, qui savons *grosso modo* le nombre de difficultés liées à la traduction automatisée et ce qui intéresse le grand public, qui croit soit que la traduction automatique, c'est pour demain, si ce n'est déjà réalisé (ou en tout cas qu'il n'y a pas de problèmes spécifiques liés à l'utilisation des ordinateurs pour traiter les langues naturelles), soit que la TA est un monstre qu'il faut combattre tous azimuts car il remplacera les secrétaires bilingues ou trilingues.

Je pense en effet que l'exemple de Winger92 illustre très bien un cas typique du manque de diffusion de l'information entre laboratoires et industrie. Le revers technologique que constitue la commercialisation de Winger92 est pourtant à considérer comme une bonne chose. Car elle contribuera à créer une prise de conscience, tant chez les utilisateurs (et il y en a), des difficultés liées à traduire, que chez les traducteurs professionnels qu'il doit y avoir une possibilité de faire mieux.

En fin de compte, elle obligera les scientifiques à faire un effort sérieux de vulgarisation.

Intégration des logiciels

La dernière décennie a vu paraître de nombreuses propositions et définitions de la station de travail du traducteur. Un trait commun à ces propositions a presque partout été l'intégration de plusieurs modules de logiciels, chacun pouvant parachever des parties bien définies de la traduction. Qu'on pense à l'article de Martin Kay : *The Proper Place of Men and Machines in Machine Translation*. Dans cet article, l'auteur avait proposé la création d'une station de travail du traducteur (*translator's amanuensis*). Celle-ci comprendrait notamment des modules de dictionnaires électroniques (destinés aux vérifications terminologiques) et des outils appropriés pour manier les textes.

Une autre tendance a été la décomposition du processus de traduction en plusieurs phases. Je ne pense pas ici à la traditionnelle triade *analyse, transfert et génération*

des systèmes automatisés, mais au fait que le processus de traduction peut être divisé en plusieurs composantes (en fonction du type de texte et des spécifications pour sa traduction), de façon que certaines phases se laissent exécuter automatiquement, d'autres semi-automatiquement ou de manière interactive, et d'autres enfin par l'homme, assisté par la machine. Le premier (ou un des tout premiers) à avoir compris cela est, à ma connaissance, Alan Melby. Dans un article paru en 1982, il propose de confier les parties standard d'un texte à traduire aux soins d'un module automatisé, alors que pour les éléments plus compliqués, on a le choix entre des sessions interactives ou bien la traduction manuelle à l'aide des dictionnaires.

Au cours des années quatre-vingt, on a vu l'avènement de plusieurs logiciels partiels qui correspondent partiellement à ces exigences ou les satisfont.

Je pense, parmi d'autres, au logiciel Reftex que j'ai conçu moi-même.

Pour ceux parmi vous qui ne connaissez pas les principes du logiciel REFTEX, je me permets de les décrire brièvement.

Il s'agit d'un logiciel dans lequel le traducteur humain se sert non pas de dictionnaires, mais de textes qui sont déjà traduits, donc présents en versions originale et traduite(s). À l'aide de ce logiciel, le traducteur peut chercher les mots ou les locutions qu'il ne connaît pas ou bien ceux qu'il ne trouve pas ou qui sont assortis d'une traduction insatisfaisante dans les dictionnaires. Le logiciel lui fournira des contextes contenant le(s) mot(s) recherché(s) ainsi que la traduction du passage parallèle dans la langue d'arrivée. Il appartient ensuite au traducteur d'évaluer les propositions reçues et de juger si elles sont pertinentes dans le cas de la traduction à faire.

Mais, à ma connaissance, on n'a pas jusqu'à ce jour vu de logiciel ou plutôt de progiciel qui intègre toutes les composantes requises. Je crois qu'on verra dans dix à quinze ans des essais d'intégration des différents logiciels. Ces logiciels vont s'appuyer sur des techniques maîtrisées, donc stables, et comprendront des analyseurs morphologiques et syntaxiques assez fiables, éventuellement assortis de modules de sémantique dite primitive.

Il s'agira probablement de logiciels visant des sous-domaines dont le lexique et la syntaxe sont faciles à circonscrire, tel l'exemple suisse de la traduction des bulletins d'avalanches, ou bien les travaux des PTT au Canada pour la mise en place d'une station de travail du traducteur.

TAO personnelle

Dans les années à venir, les besoins de traduction ne cesseront d'augmenter si bien que les retards connus dès aujourd'hui iront en s'accumulant. La raison en est notamment l'internationalisation accrue de toutes sortes de relations commerciales, scientifiques, culturelles, etc. Les demandes accrues de traductions que les traducteurs professionnels ne sauront pas satisfaire, auront des impacts importants sur le processus de traduction. Ainsi, on verra que les spécialistes non linguistes, qu'ils soient économistes, ingénieurs ou chimistes, n'accepteront plus les délais imposés par les services de traduction. L'avènement de l'Europe unique aura également pour conséquence que

de nombreuses PME, qui ont basé jusqu'ici leurs activités exclusivement sur le marché intérieur, essentiellement monolingue (l'exception notoire étant la Belgique), auront besoin de communiquer en langue(s) étrangère(s).

Ces deux tendances font qu'on verra évoluer des logiciels qui permettront aux spécialistes de faire leurs propres traductions.

Une pareille évolution exigera une redéfinition du rôle des traducteurs. Leur rôle, jusqu'à maintenant interprétatif-simultané, évoluera vers un statut de « confirmateur »-« réviseur ». On n'a qu'à prendre l'exemple des négociations communautaires : les hommes politiques n'attendent pas que les traductions soient peaufinées, ils essaient de communiquer et de parvenir à des accords, tant bien que mal.

Mais cette tendance signifiera également une révolution dans la façon de penser et de concevoir les auxiliaires de traduction assistée par ordinateur (TAO). Pour permettre aux spécialistes non linguistes de traduire et de communiquer dans une langue étrangère, il leur faut des outils qui pourront les aider. On verra donc des logiciels travaillant en sens inverse, c'est-à-dire qui traduisent depuis la langue maternelle vers la langue étrangère — jusqu'à présent, la norme a été de traduire vers la langue maternelle.

À ma connaissance, il existe deux projets qui vont dans ce sens. L'un est le projet LIDIA, sous la direction de Christian Boitet, du GETA à Grenoble. L'autre est plutôt un avant-projet, baptisé MITRA, visant justement les PME, et dont l'idée vient d'une équipe de l'École des hautes études commerciales d'Aarhus au Danemark.

Bibliographie

BOITET C. (1990) « Vers la TAO personnelle : Le projet LIDIA du GETA », *La Tribune des Industries de la Langue*, juillet-octobre 90.

HOLMBOE H. (1989) *MITRA*, École des hautes études commerciales d'Aarhus au Danemark. (document interne).

KAY M. (1980) *The Proper Place of Men and Machines in Machine Translation*, Palo Alto (Cal.).

KJÆRSGAARD P. S. (1987) « REFTEX — A Context-Based Translation Aid », *Proceedings of Third Conference of the European Chapter of the Association for Computational Linguistics*, Morristown.

KJÆRSGAARD P. S. (1989) « REFTEX — Un progiciel pour la traduction assistée par ordinateur », *Meta*, Vol. 34, n° 3.

MELBY A. K. (1981) « Translators and Machines — Can They Cooperate? », *L'informatique au service de la traduction, Meta* 26, n° 1.

WNGER92 : Rapport n° 2 om maskinel sprog-oversættelse, document interne 1989 (en français : rapport n° 2 sur la TAO)

WINGER92 : PC-baseret oversættelse med Winger92, document interne 1989 (en français : traduction sur PC à l'aide de Winger92).

18

Un projet de dictionnaire intelligent

Jean-Marie MAES

Provinciaal Instituut voor Hoger Onderwijs (PIHO), Gand, Belgique

Résumé

ID (*intelligent dictionary*) est un outil électronique puissant qui a été développé pour les cours de traduction, mais qui pourrait être très utile au professionnel aussi. ID étend largement les possibilités d'accès aux données en permettant l'utilisation de trois clés ou une combinaison quelconque de ces clés (l'entrée, le code sémantique et le code grammatical). Les fiches du dictionnaire sont automatiquement liées aux contextes originaux, permettant ainsi à l'utilisateur d'explorer les textes sources et les textes cibles déjà traduits à partir de la fiche. Il n'y a pratiquement pas de limites à l'information qu'une fiche peut contenir, étant donné que chacune peut être liée à un texte ASCII externe qui peut être consulté et exploité à partir du programme. L'utilisateur peut établir librement son circuit de références.

La version résidente d'ID n'a besoin que de 8 Ko et offre toute la puissance de la partie dictionnaire du programme et en plus une fonction *cut & paste* (une fonction de découpe et d'insertion).

ID est équipé d'une fonction d'aide directement appliquée au contexte.

Toutes ces caractéristiques font qu'ID est devenu un système de récupération d'information, basé sur le lexique (*a lexicon-based information retrieval system*), plutôt qu'un simple dictionnaire électronique.

Introduction

Les systèmes de dictionnaires électroniques *versus* ID

On peut distinguer deux systèmes de dictionnaires électroniques : des systèmes clos utilisant CD-ROM et des systèmes ouverts permettant à l'utilisateur de compiler ses dictionnaires personnels.

Dans un système clos, on peut seulement consulter les données. Il n'est pas possible de changer ni d'ajouter quoi que ce soit.

Dans un système ouvert comme ID, l'utilisateur doit compiler lui-même ses données lexicographiques. Dans le stade initial, cela peut être fastidieux, mais ces dictionnaires personnalisés sont extrêmement utiles au traducteur. Beaucoup de traducteurs se spécialisent dans certains sujets. Un système ouvert leur permet de compiler des dictionnaires spécifiques pour chacun des sujets. Ces dictionnaires se révèlent souvent plus utiles que les dictionnaires classiques généraux et spécialisés, puisqu'ils sont tout à fait orientés vers les besoins personnels du traducteur.

La plupart des systèmes ouverts de dictionnaire électronique offrent un accès très limité aux dictionnaires, le plus souvent seulement par l'entrée. Cela est dommage, puisque cette restriction n'a rien à voir avec le matériel (*hardware*).

Le traducteur utilise des dictionnaires bilingues, mais aussi des dictionnaires explicatifs et des lexiques systématiques. Ce serait commode si l'on réussissait à intégrer ces trois systèmes de compilation et à les rendre accessibles à partir d'un programme de traitement de texte.

Si l'on savait intégrer la structure sémantique du lexique dans le dictionnaire électronique, il serait possible d'y accéder à partir d'un seul fichier de données.

C'est précisement ce qui se fait dans ID.

ID ne construit qu'un seul ensemble de données, mais l'utilisateur peut y accéder en utilisant trois clés ou une combinaison quelconque de ces clés :

- l'entrée;
- le code sémantique;
- le code grammatical.

Chacune de ces clés peut être incomplète ou vide, mais elles ne peuvent pas être vides toutes les trois. Ainsi l'utilisateur a plusieurs possibilités d'accès aux données. Il peut changer la clé active à chaque instant. Il n'y a pas de limite au nombre de fois qu'une clé peut être utilisée. Seule la combinaison des trois clés doit être unique (ce qui est logique, ces clés définissant un concept). Ainsi ID intègre un dictionnaire explicatif, un dictionnaire bilingue et un lexique organisé systématiquement. Les trois clés produisent un effet de zoom très flexible, permettant à l'utilisateur d'accéder aux données selon sa préférence.

ID ne se limite pas à retrouver des termes dans un dictionnaire. Le programme couvre tout le processus de traduction à partir du texte source, par le dictionnaire, jusqu'au texte cible.

L'organigramme du programme peut être illustré ainsi (fig. 1) :

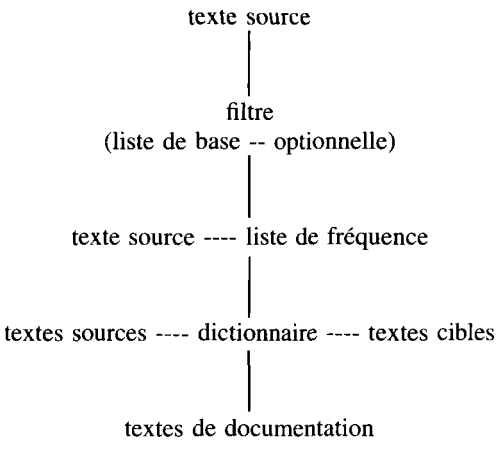

texte source
filtre
(liste de base -- optionnelle)
texte source ---- liste de fréquence
textes sources ---- dictionnaire ---- textes cibles
textes de documentation

Figure 1

Le programme ayant d'abord été développé pour les cours de traduction, il devait être très facile à utiliser. Toutes les fonctions du dictionnaire sont intégrées, ce qui évite le lent cheminement à travers de longs menus. Ainsi ID permet une utilisation rapide et commode. En plus il est équipé d'une fonction d'aide avancée, appropriée au contexte, avec des références croisées, un index et de l'aide sur la fonction d'aide (*help on help*).

ID : le programme

ID est un ensemble de programmes complet, offrant beaucoup plus que l'accès au dictionnaire. ID peut être utilisé indépendamment de la langue : on peut compiler des dictionnaires des principales langues occidentales. ID a été testé pour le français, le néerlandais, l'allemand et l'anglais. Tous les caractères spéciaux de ces langues (et d'autres) peuvent être utilisés sans problème (par exemple : le programme distingue « ê » de « e », et il les classe par ordre alphabétique correct.)

ID maintient le lien entre les textes et le dictionnaire à tous les stades du développement d'un dictionnaire. Ainsi l'utilisateur peut toujours retourner aux contextes originaux des textes sources.

Du texte source à la liste de fréquence

À partir d'un texte ASCII quelconque, ID construit une liste de fréquence contenant toutes les formes lexicales du texte. Cette liste peut être filtrée. Toutes les formes qui se trouvent dans une liste de base sont alors omises dans la liste de fréquence.

Le nombre de listes de base est illimité.

(Dans un contexte didactique, le professeur peut utiliser des listes de base différentes, adaptées au niveau des étudiants).

L'utilisateur peut éditer les formes d'une manière aisée et rapide. Il peut retourner au texte source quand il veut. Ceci l'aide à détecter les formes composées.

Voici, à titre d'illustration, un écran montrant une partie d'une liste de fréquence d'un texte anglais (*cf.* fig 2).

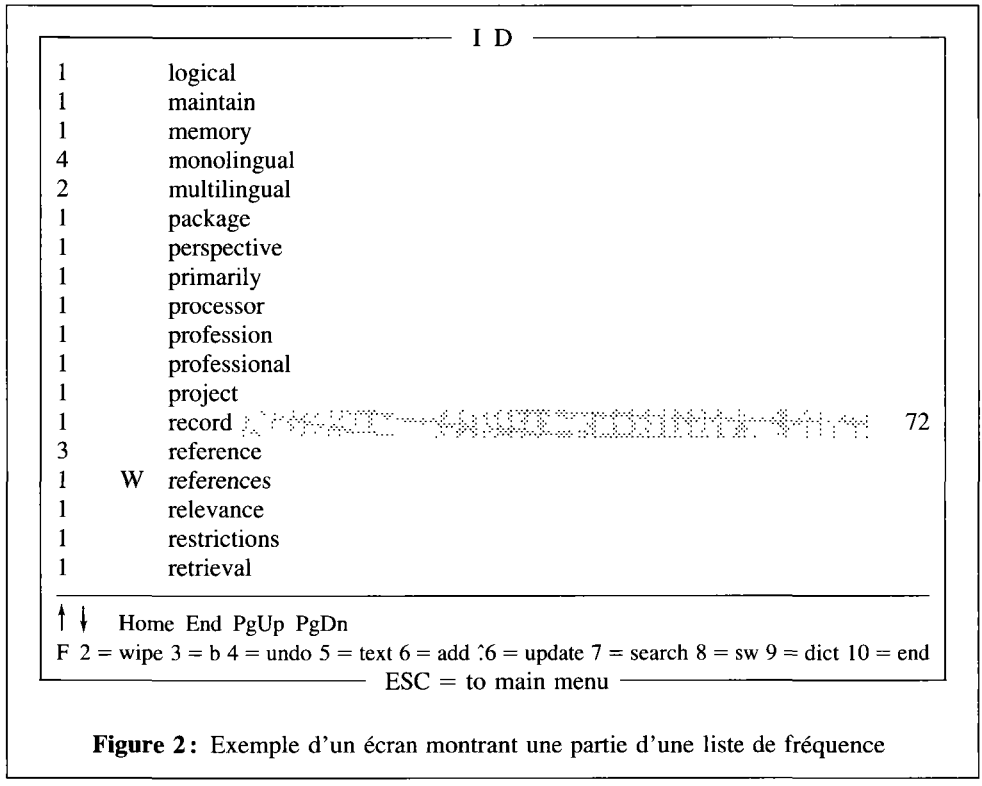


Figure 2: Exemple d'un écran montrant une partie d'une liste de fréquence

L'utilisateur peut se déplacer en utilisant les touches de curseur. Il peut éditer chaque ligne avec les commandes traditionnelles de traitement de texte. Les touches de fonction ont une signification spécifique :

- F2 est utilisée pour effacer une forme.
- F3 efface une forme et la transfère à la liste de base sélectionnée. Cette manière d'enrichir la liste de base en travaillant les listes de fréquence est particulièrement commode.
- F4 annule les actions de F2 et F3.
- F6 et ^F6 sont utilisées pour ajouter des formes à la liste et pour la mettre à jour.

- F7 permet à l'utilisateur d'aller vers une ligne ou un mot particulier.
- F8 vous conduit au champ des fréquences.
- F9 termine la session de travail et transfère toutes les formes résultantes au dictionnaire choisi.
- F10 termine la session de travail. L'utilisateur peut évidemment reprendre son travail plus tard.
- F5 vous conduit au texte source. La première occurrence du mot (ou du radical du mot) est marquée. L'utilisateur peut se déplacer vers l'occurrence suivante. Il dispose de toutes les fonctions traditionnelles d'un programme de balayage (*browser*).

Dans notre exemple, il peut aller du mot **record** dans la liste de fréquence à la première occurrence du mot dans le texte source (cf. fig 3).

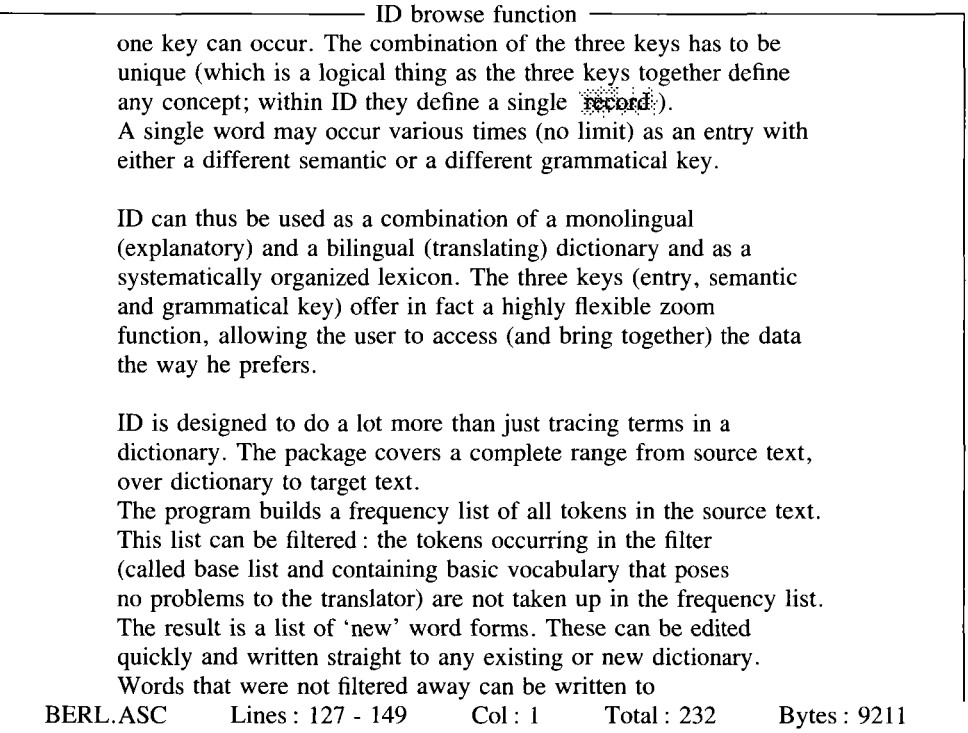

ID browse function
one key can occur. The combination of the three keys has to be
unique (which is a logical thing as the three keys together define
any concept; within ID they define a single record).
A single word may occur various times (no limit) as an entry with
either a different semantic or a different grammatical key.
ID can thus be used as a combination of a monolingual
(explanatory) and a bilingual (translating) dictionary and as a
systematically organized lexicon. The three keys (entry, semantic
and grammatical key) offer in fact a highly flexible zoom
function, allowing the user to access (and bring together) the data
the way he prefers.
ID is designed to do a lot more than just tracing terms in a
dictionary. The package covers a complete range from source text,
over dictionary to target text.
The program builds a frequency list of all tokens in the source text.
This list can be filtered : the tokens occurring in the filter
(called base list and containing basic vocabulary that poses
no problems to the translator) are not taken up in the frequency list.
The result is a list of 'new' word forms. These can be edited
quickly and written straight to any existing or new dictionary.
Words that were not filtered away can be written to
BERL.ASC Lines : 127 - 149 Col : 1 Total : 232 Bytes : 9211

Figure 3 : Exemple d'un écran montrant une partie du texte source

Cette possibilité de retourner rapidement aux contextes originaux du texte source (en tenant compte de la manière dont les formes déclinées et conjuguées ont été changées pour former l'entrée) est un des atouts les plus typiques et importants du programme.

Il est possible d'envoyer une liste de fréquence à l'imprimeur ou de la transférer à un texte ASCII (pour l'utiliser dans un programme de traitement de texte par exemple).

Le filtre

Un filtre (ou liste de base) est une liste d'occurrences (de formes lexicales) qu'on ne veut PAS sélectionner pour la liste de fréquence et le dictionnaire. L'utilisateur peut éditer une liste de base à peu près comme une liste de fréquence. Il peut y ajouter des formes, éditer les formes, chercher une forme particulière, etc.

Une liste peut être imprimée ou transférée à un texte ASCII. Il n'y a pas de limite au nombre de filtres possibles. (Dans la version 1.01 d'ID le nombre d'entrées d'une liste est limité à 15 000. En réalité, le nombre n'est pas limité, mais s'il excède 15 000, la liste ne peut pas être travaillée à l'écran.)

Le dictionnaire

Quand on a édité une liste de fréquence, on l'envoie normalement vers le dictionnaire de son choix. Si le dictionnaire existe déjà, seules les entrées nouvelles lui sont ajoutées. Le champ des contextes est automatiquement mis à jour pour les entrées déjà présentes, comme pour les entrées nouvelles. Ceci permet à l'utilisateur de retourner au texte source de son choix à partir de la fiche du dictionnaire.

Les entrées nouvelles ont *dummy* comme code sémantique et grammatical. Cela permet une édition systématique des entrées nouvelles (en les cherchant avec *dummy* comme code sémantique). ID se rappelle le code sémantique et grammatical que l'on utilise pour consulter le dictionnaire, même si l'on a changé ces codes.

Ayant choisi un dictionnaire, l'utilisateur peut entreprendre trois actions :

- AJOUTER une entrée;
- CHERCHER une entrée spécifique;
- utiliser les fonctions GÉNÉRALES.

Ajouter

L'utilisateur se retrouve dans une fiche nouvelle. La date est générée automatiquement. Il peut éditer tous les champs de la fiche comme il veut, mais il doit toujours remplir les trois champs clés puisqu'ils définissent une fiche unique :

- le champ de l'entrée;
- le champ du code sémantique;
- le champ du code grammatical.

Il peut se servir des listes des codes sémantiques et grammaticaux utilisés pour faire une sélection.

Chercher

On peut rechercher une fiche quelconque en utilisant les champs clés : on entre une combinaison quelconque de l'entrée et des codes sémantiques et grammaticaux.

Chacune de ces clés peut être incomplète ou vide, mais elles ne peuvent pas être vides toutes les trois. Cette combinaison des clés détermine le mode d'accès au dictionnaire : elle précise la manière dont l'utilisateur peut parcourir les données du dictionnaire. S'il a sélectionné un code sémantique particulier, il travaillera avec ce sous-ensemble particulier des données.

Ce qui est très intéressant, c'est que l'on peut changer à chaque instant la combinaison des clés actives.

Dans le dictionnaire de test que j'utilise, l'entrée incomplète « id » résultera dans l'écran suivant (fig. 4) :

```
ID   D:\FID1\TESTE-D                                        Nr of rec. : 28

ENTRY : id project                                                12 06 1990
        SEM. KEY : cptr ai              GRAM. KEY : n-count-prop

        Translat-1 : id-project
        Translat-2 :
        Translat-3 :
Definition :
Project for the development of an 'intelligent dictionary' (ID) system, in which the
functions of monolingual and bilingual dictionaries and of a « lexicon » are combined.
The dictionary 'remembers' the source texts the entries come from.
This enables the user to return to the original contexts of a term and even to consult
already translated texts.
In this way ID has grown into a lexicon steered information retrieval system.
There are plans to further develop the program as an information retrieval system and
for use in semi-automatic dictionary building.
Contacts are being laid with institutes willing to cooperate.
See : « electronic dictionary ».
EncY : IDTXT.ASC                          Cont : IDMEMO
Sources :  T1 :        T2 :        T3 :        D : jmm        C :

Search : N)ext, P)rev, K)ey, E)dit, W)ipe, Q)uit?
Search key = ENTRY
```

Figure 4 : Exemple d'une entrée de dictionnaire

Une fiche peut contenir beaucoup d'information. Des sources peuvent être indiquées pour les traductions et la définition. Il est souvent essentiel d'avoir accès à une information plus vaste, en particulier dans le cas des termes techniques et scientifiques.

C'est pourquoi ID offre la possibilité de lier **chaque** fiche à un texte ASCII externe. Le champ marqué *EncY* est utilisé à cet effet. Ainsi l'utilisateur peut balayer ce texte externe à partir de la fiche.

Le champ des contextes (*Cont*) contient le nom de trente textes sources maximum par fiche, ce qui permet de retourner à tous les contextes originaux de chacun de ces

textes. De même on peut retourner aux textes cibles des textes déjà traduits. On peut ainsi comparer les traductions possibles dans leur contextes originaux. **De cette manière l'utilisateur a toute son expérience de traduction sous la main, moyennant quelques touches.**

À partir d'une fiche, l'utilisateur peut entreprendre des actions diverses :

- **N**)ext (ou flèche vers le bas) fait apparaître l'entrée suivante. Le résultat depend de la clé active (ou de la combinaision particulière des clés actives). Inutile de retenir cela, puisque la clé active est marquée au bas de l'écran. Dans notre exemple la clé active est l'entrée. Cela veut dire que la fiche suivante sera l'entrée suivante dans l'ordre alphabétique, indépendamment des catégories sémantiques et grammaticales. Si l'on avait sélectionné une clé sémantique et une clé grammaticale, la fiche suivante serait l'entrée suivante dans les (sous-)catégories sémantiques et grammaticales spécifiées.
- **P**)rev (ou flèche vers le haut) donne l'entrée précédente. Le résultat dépend de nouveau des clés actives.
- **K**)ey permet à l'utilisateur de changer la clé active ou la combinaison de ces clés. Ceci lui confère une grande flexibilité d'accès aux données, même pendant la consultation.
- **E**)dit permet d'éditer les champs de la fiche. On peut utiliser les possibilités traditionnelles d'édition (même le *word wrapping* dans le champ de la définition : pas de découpe en fin de ligne).
- **W**)ipe efface la fiche et fait revenir l'utilisateur au niveau précédent.
- On peut aller directement aux champs qui sont marqués d'un caractère éclairé (*highlighted*) : « 1 », « 2 », « 3 » vous conduisent au champ de traduction correspondant : « D » vous conduit au champ de définition, « C » au champ des contextes et « Y » au champ du texte externe.

Il est possible de consulter une liste des codes utilisés à partir du champ sémantique et grammatical (cf. fig. 5).

On utilise les touches de curseur ou l'on entre les premiers caractères d'un code pour le sélectionner. Si l'on entre un code nouveau, la liste des codes est mise à jour automatiquement.

Il est très facile de créer des références partout dans le champ de définition. Il faut seulement marquer le début et la fin de la référence (en utilisant les touches de fonction F2 et F3 respectivement). La touche de fonction F5 vous conduit à la fiche de référence. Si cette fiche n'existe pas, elle est créée. On bénéficie toujours de toutes les fonctions du dictionnaire. Il n'y pas de limite au nombre des références que l'on peut établir. En plus, on peut voyager de référence à référence sans risque de se perdre : on peut toujours retourner à la fiche de départ d'un seul coup. La seule limite, c'est l'espace du champ de définition.

Ce système de référence permet à l'utilisateur de créer et d'exploiter des voies particulières dans la banque de données.

```
ID  D:\FID1\TESTE-D                                          Nr of rec. : 28

ENTRY : id project                                                12 06 1989
     SEM. KEY : cptr ai                    GRAM. KEY : n-count-prop

   Translat-1 : id-project
   Translat-2 :  ___ SEMANTIC CODES ___
   Translat-3 : | cptr                  |
Definition :    | cptr ai              |
Project for the | gen                  |  elligent dictionary' (ID) system, in
which the funct | gen mach             |  bilingual dictionaries and of a lexicon
are combined.   | gen class            |
The dictionary  | gen psych            |  texts the entries come from.
This enables th | ling                 |  original contexts of a term and even to
consult already | ling hist            |
In this way ID  |       NEW CODE       |  n steered information retrieval system.
There are plans |_____      ↑ more ____|  program as an information retrieval
system and for use in semi-automatic dictionary building.
Contacts are being laid with institutes willing to cooperate.
See : « electronic dictionary ».
EncY : IDTXT.ASC                                  Cont : IDMEMO
Sources :  T1 :          T2 :          T3 :          D : jmm          C :

Search : N)ext, P)rev, K)ey, E)dit, W)ipe, Q)uit?
Search key = ENTRY
```

Figure 5 : Exemple d'une entrée avec les codes sémantiques

Fonctions générales

Ces fonctions offrent à l'utilisateur un aperçu du contenu du dictionnaire ou d'une partie du dictionnaire. Il peut de nouveau utiliser les trois clés pour s'adresser au dictionnaire. Un exemple : s'il veut savoir quels substantifs appartenant au champ lexical de l'ordinateur se trouvent dans le dictionnaire, il lui suffit de séléctionner les clés en question. Le résultat est une liste d'entrées avec leurs codes sémantiques et grammaticaux correspondant à ces critères (fig. 6).

Pour aller à la fiche elle-même, il suffit d'entrer la lettre à gauche de l'entrée. On peut utiliser toutes les fonctions du dictionnaire. Il est donc possible d'aller de la liste à des fiches particulières du dictionnaire, d'aller ensuite de la fiche vers les textes sources et (ou) les textes cibles et puis de retourner à la liste générale. En outre, tout cela peut se faire à partir d'un programme de traitement de texte.

Il n'est pas seulement possible de consulter cette information générale à l'écran, on peut aussi envoyer le contenu partiel ou complet des fiches à l'imprimeur ou le transférer à un texte ASCII.

ID	D:\FID1\TESTE-D		Nr of rec. : 28
a	access	cptr	n c
b	network	cptr	n c
c	record	cptr	n c
d	frame	cptr ai	n c
e	id project	cptr ai	n c prop
f	lemmatizer	cptr ai	n c
g	net	cptr ai	n c
h	semantic net	cptr ai	n c
...			

Strike <RETURN>

Figure 6: Exemple d'une liste d'entrées aves leur clés

IDSWAP et les programmes auxiliaires

IDSWAP et IDRES sont des programmes résidents. Ils permettent d'utiliser des dictionnaires au choix à partir d'un programme de traitement de texte (non graphique). On peut aller immédiatement d'un mot du texte vers la fiche correspondante ou aller vers le premier niveau du programme (ajouter, chercher, fonctions générales). Il est possible d'insérer (*paste*) la traduction de votre choix dans le texte sous-jacent ou de la transfèrer vers un tampon et de l'insérer plus tard. Il est possible de changer rapidement de dictionnaire. (Les dictionnaires peuvent se trouver partout sur un ou plusieurs disques durs.)

IDSWAP offre toutes les fonctions de la partie dictionnaire d'ID (IDRES n'a pas de fonction d'aide et de balayage des textes externes).

ID offre en plus quelques programmes auxiliaires groupés dans IDUTIL. Une des possibilités les plus intéressantes vous permet d'unir deux dictionnaires ou de transférer une partie d'un dictionnaire (sélectionnée avec les trois clés disponibles) vers un autre dictionnaire (existant ou nouveau). Il est aussi possible d'éditer d'un seul coup les clés sémantiques et grammaticales d'un dictionnaire pour toutes les fiches.

Conclusions et projets

Je n'ai pu vous présenter qu'un aperçu général du programme. Il y a beaucoup plus que je n'ai pu mentionner ici. Il est clair cependant qu'ID est un outil puissant pour chacun dans le domaine de la traduction et même des langues en général.

Les caractéristiques les plus importantes du programme sont :

- la flexibilité d'accès aux données d'une manière intelligente en utilisant l'entrée, le code sémantique ou grammatical ou une combinaison quelconque de ceux-ci;

- l'intégration des fonctions du dictionnaire permettant d'effectuer la plupart des fonctions au même niveau du programme (même de changer les clés actives pendant que l'on se trouve dans le dictionnaire);
- la possibilité de retourner aux contextes originaux des textes sources et de les comparer aux textes cibles déjà traduits;
- le système de référence offrant des possibilités presque illimitées;
- la simplicité d'usage et la structure logique du programme.

Quant au matériel, un simple PC XT compatible avec disque dur suffit. Un AT avec un disque dur à grande capacité est la plate-forme idéale pour ID.

Nous avons des projets tendant à développer ID encore plus comme un système d'information basé sur le lexique. (Une version d'ID pour réseau est prévue pour la fin de l'année 1991.) Les expériences des utilisateurs d'ID peuvent nous être très utiles dans ce développement.

19

EURODICAUTOM, ou la terminologie en l'an 2001

Alain REICHLING

Commission des Communautés européennes, Service de traduction, Luxembourg

Contexte de développement

EURODICAUTOM est la banque de données terminologiques multilingues développée au sein du Service de traduction de la Commission des Communautés européennes. Ce grand fichier terminologique automatisé donne des informations terminologiques (termes, synonymes, définitions, contextes, notes) sur plus de 500 000 concepts différents, ce qui représente près de trois millions de termes, et contient plus de 150 000 abréviations, sigles ou acronymes. Il couvre les différents domaines d'activité de la Commission, et cela dans les neuf langues officielles et de travail de la Communauté.

La décision de créer cet outil informatique est la conséquence logique de la volonté politique des pères de l'Europe de traiter sur un pied d'égalité toutes les langues officielles des États membres, décision confirmée à chaque élargissement de la Communauté par le Conseil des ministres.

En effet, les traducteurs ont été confrontés dès le début, et sans cesse davantage au fur et à mesure que les initiatives de la Commission touchaient des domaines de plus en plus variés et que le nombre de langues augmentait (pour représenter aujourd'hui 72 combinaisons linguistiques), à une multitude de problèmes terminologiques qu'aucun d'eux ne pouvait maîtriser totalement. Nul ne peut prétendre maîtriser, en dépit du haut niveau de ses connaissances linguistiques et de sa relative spécialisation dans un ou plusieurs domaines de travail, la totalité des terminologies qui existent dans

sa langue maternelle et dans les quatre ou cinq langues à partir desquelles il est régulièrement appelé à travailler.

L'informatique, et plus particulièrement le développement d'une banque de données terminologiques, est très tôt apparue comme la seule réponse à ce défi quotidien, face auquel des outils traditionnels comme les dictionnaires, les glossaires, les fichiers paraissaient insuffisants. En effet, un ordinateur permet de maintenir à jour les contenus et d'organiser l'accès aux données d'une façon beaucoup plus efficace.

À la Commission des Communautés européennes, on peut dire qu'EURODICAUTOM est désormais considéré par les traducteurs comme un outil de travail tout à fait banalisé. Par contre, une étude récente effectuée par une étudiante de l'Université de Mons, confirmée par d'autres enquêtes d'ailleurs, montre que l'informatique, et plus particulièrement la terminologie informatisée, est encore une denrée rare dans l'environnement de travail du traducteur indépendant. Les organisateurs de ce Colloque ont donc pensé qu'une présentation d'EURODICAUTOM, tel qu'il existe et est utilisé aujourd'hui, intéresserait ceux pour qui un tel outil ne deviendra une réalité que demain ou après-demain, c'est-à-dire en l'an 2001.

Objectifs d'EURODICAUTOM

Les principaux objectifs visés par l'intégration d'EURODICAUTOM dans l'environnement de travail du traducteur sont liés à trois soucis constants : augmenter la productivité, accroître la qualité, faciliter l'harmonisation de la terminologie utilisée dans les textes. Il n'y a pas grand risque à affirmer que ces objectifs ne sont pas propres aux services de la Commission et qu'ils continueront à guider les développements futurs.

Les traducteurs eux-mêmes s'accordent à considérer que la recherche terminologique entre pour 20 à 70 pour cent dans le travail de traduction. Tout gain de temps dans la quête du mot juste contribue donc directement à l'augmentation du nombre de pages traduites, ou rend au traducteur un peu de cette denrée rare qu'est le temps, si nécessaire pour peaufiner le style d'un texte traduit. Or il est évident que l'accès direct, à partir de son bureau, à une multitude d'informations terminologiques habituellement éparpillées en différents endroits parfois difficilement accessibles, et la possibilité pour tous les utilisateurs d'un système de prendre connaissance instantanément des trouvailles terminologiques d'un collègue isolé, sont des facteurs qui réduisent de beaucoup le temps de recherche. D'ailleurs, même lorsque la machine ne trouve pas de réponse, il en résulte un gain de temps appréciable, puisque l'utilisateur sait, après une demi-seconde, que la réponse à sa question ne se trouve pas dans les sources qu'il consulte habituellement : il pourra donc organiser une recherche plus poussée sans perdre de temps.

La qualité terminologique des textes bénéficie également de l'utilisation d'un système central de terminologie, car le traducteur peut confronter les solutions proposées par différents collaborateurs, aussi bien documentées que possible, de sorte qu'il peut exercer son sens critique habituel avec une pertinence accrue; par ailleurs, les com-

mentaires d'utilisateurs compétents contribuent à améliorer, à compléter, à corriger le contenu de la base.

Le recours à une même source d'information contribue en outre dans une large mesure à l'harmonisation de la terminologie dans les documents communautaires, et, partant, à l'amélioration de la communication scientifique et technique en Europe.

Caractéristiques d'EURODICAUTOM

Lors de la conception du système EURODICAUTOM, un certain nombre de contraintes ont dû être prises en considération. Les solutions imaginées alors restent valables aujourd'hui et méritent d'être développées davantage.

Le premier critère auquel doit satisfaire un système informatique destiné non pas à un informaticien ou à un chercheur, mais à un praticien, c'est certainement la simplicité d'utilisation, la convivialité. Il faut en quelque sorte que l'utilisateur oublie qu'il a un système informatique devant lui. Dans le cas de la terminologie dans un contexte traductionnel, ceci signifie que l'utilisateur final n'a qu'à taper le terme, simple ou complexe, sur le clavier de son PC, et la machine doit s'occuper du reste.

Un système informatique doit également être souple et dynamique. Souple, pour accepter différents types de données, des mots simples, des termes complexes, des éléments phraséologiques, des contextes en tout genre, etc. Souple aussi pour que le traitement de ces données n'exige pas un effort de préparation excessif, mais que des algorithmes appropriés indexent correctement les documents : les ressources humaines sont rares et chères. Souple encore, pour permettre une utilisation à partir de postes de travail très diversifiés, même les moins intelligents, donnant un accès à partir de n'importe quel point d'un réseau.

Le dynamisme du système doit permettre une évolution aisée du logiciel, en fonction d'une part des développements technologiques, d'autre part, de la naissance ou de la reconnaissance de nouveaux besoins. Ainsi, dans le cas de la Commission, il a fallu faire évoluer le progiciel EURODICAUTOM pour prendre en compte des langues supplémentaires à la suite des élargissements successifs de la Communauté.

La caractéristique essentielle toutefois d'un système automatisé doit être son efficacité. Peu importe qu'il puisse faire beaucoup de choses, s'il ne peut pas faire correctement celles que l'on attend de lui. Dans le cas d'EURODICAUTOM, le but premier est de donner instantanément au traducteur la réponse à un problème terminologique rencontré au cours de la traduction d'un texte, et c'est ce que fait le système.

EURODICAUTOM propose donc au traducteur la réponse au problème terminologique qui lui est soumis, du moins s'il la trouve dans sa mémoire. C'est là le minimum que l'on est en droit d'attendre d'une banque de données terminologiques. EURODICAUTOM se distingue toutefois d'autres systèmes par la philosophie qui a guidé sa mise au point et qui a inspiré d'autres chercheurs depuis. En effet, conscients de la difficulté qu'il y a à recenser toutes les terminologies nécessaires, les concepteurs de la base ont rejeté une approche du type « tout ou rien ». Ils ont préféré des solutions

qui n'établissent pas d'exclusive par rapport à l'information terminologique (ou phraséologique) à fournir.

Le système prévoit donc que les données les plus pertinentes apparaissent en tête des réponses possibles, mais que des réponses incomplètes ou seulement partiellement satisfaisantes ne soient pas écartées. À défaut d'une réponse correspondant totalement aux paramètres de la recherche, il proposera, par exemple, une réponse dans une langue autre que celle demandée, ou le terme au pluriel, ou un élément de syntagme, ou une donnée partielle dans un autre domaine, etc. L'efficacité du système dans un environnement traductionnel s'en trouve ainsi accrue.

Par souci de convivialité, tout ceci doit se faire sans intervention du traducteur ou du terminologue, sans donc que l'utilisateur doive faire plus que taper sa question, sans qu'il lui faille spécifier des opérateurs de recherche particuliers.

Le but recherché est obtenu par la mise en œuvre d'algorithmes appropriés, tant au niveau des procédures d'indexage des données qu'au niveau des stratégies de recherche. On pourrait comparer EURODICAUTOM à une voiture à boîte de vitesses automatique, ou à un appareil photographique à mise au point automatique, ce qui montre que le système, bien qu'en place depuis plusieurs années déjà, n'a rien perdu de sa modernité.

Les aspects concernant plus spécifiquement les possibilités de consultation en différé et la gestion de la terminologie ne seront pas examinés ici.

Développements prévisibles

EURODICAUTOM est appelé à s'intégrer de plus en plus dans le poste de travail informatisé du traducteur et sera complété par des outils de repérage, de production et de gestion de la terminologie. Il pourra constituer un maillon important des outils de rationalisation des textes et de nouveaux systèmes de TAO, voire constituer une solution de rechange pour ceux-ci.

On peut également prévoir une interconnexion poussée et transparente pour l'utilisateur entre EURODICAUTOM et d'autres applications terminologiques (individuelles, au niveau d'une communauté d'utilisateurs, extérieures à la Commission, etc.). Des fichiers auxiliaires et des bases satellites graviteront autour du noyau central.

Pour faciliter la communication entre le traducteur et le système informatique, des systèmes de reconnaissance ou de synthèse vocale pourront être utilisés. Les termes questions pourront être repérés automatiquement dans le document source, de sorte que l'utilisateur n'ait même plus à s'occuper de la recherche proprement dite.

En amont de l'application, il s'agit notamment de diversifier et d'automatiser les processus d'extraction de termes, de développer et d'affiner les possibilités qui existent déjà de traiter automatiquement des listes de termes pour fournir au traducteur des « glossaires » synoptiques adaptés au document à traduire. Une telle « dynamisation » de la recherche terminologique est utile aussi bien au terminologue qu'au traducteur car elle contribue à l'efficacité de leur travail et à l'harmonisation systématisée des textes.

En aval, il faut diversifier les produits extraits de la base en fonction de l'évolution de la clientèle et des supports. En particulier, il faut noter que la plupart des possibilités de consultation en différé, solutions de rechange et compléments de la consultation en-ligne auquel il est fait référence plus haut, sont encore orientées vers la sortie de listes imprimées. Dès à présent, il faut prévoir la récupération de fichiers éditables permettant le (re)traitement électronique de l'information.

Au niveau de l'application elle-même, le progiciel pourra être développé dans la ligne de la philosophie qui lui est sous-jacente. Le souci d'uninotionalité qui caractérise une application terminologique sera réconcilié avec le désir de préserver l'intégrité de certaines collections terminologiques particulières : terminologies normalisées ou nomenclatures officielles, par exemple.

L'orientation « conceptuelle » sera maintenue et développée. En particulier, les possibilités actuelles de remonter de la description du concept vers le terme pourraient trouver une extension pour venir en aide aux rédacteurs de textes.

Pour renforcer encore la convivialité du système sans nuire à son efficacité, les possibilités de lier les stratégies de recherche et les modes de présentation à différents profils d'utilisateur doivent être développées.

On peut penser que l'approche automatique suivie jusqu'ici continuera à se développer, mais que, pour poursuivre ma comparaison avec l'automobile ou la photographie, il sera possible également de changer de vitesse manuellement, ou de déconnecter le mécanisme de mise au point automatique pour faire des réglages manuels. L'utilisateur qui le souhaiterait pourra donc, plus largement qu'aujourd'hui, définir lui-même les paramètres qu'il souhaite voir privilégiés. Des comparaisons chiffrées montrent toutefois que, dans la plupart des cas, l'approche automatique donne des résultats plus efficaces : ce sont des raisons essentiellement psychologiques qui justifient le maintien d'approches différentes.

C'est donc une approche « à la carte » qui se généralisera, aussi bien pour les stratégies de recherche, voire dans certains cas pour les procédures d'indexage, que pour les modes d'affichage des données, à côté des processus « automatiques » utilisant des menus à différents niveaux. L'utilisateur pourra réduire ou au contraire élargir les possibilités d'obtenir des informations partielles, en intervenant sur les combinaisons de codes matières, sur la position et la longueur de la troncature, etc.

L'utilisation d'interfaces graphiques pourra remplacer ou compléter l'utilisation de définitions. La connexion avec des bases de données encyclopédiques et des systèmes audiovisuels améliorera les possibilités de formation. L'utilisation de systèmes experts ou de bases de connaissances pourra pallier les aléas des systèmes de classification : ceci me paraît être la seule piste prometteuse dans l'utilisation efficace de réseaux notionnels, notamment pour affiner la recherche thématique des informations.

Conclusion

Il est difficile de prédire l'avenir. Au cours des dernières années, partisans des développements centraux et fanatiques des « banquettes » mises en réseau ont à tour de rôle

prédit la mort de l'une ou l'autre approche. Défenseurs du documentaire et inconditionnels du relationnel se sont opposés régulièrement. Selon les contraintes et les progrès technologiques, selon les ressources disponibles, selon les modes d'organisation et les budgets, des solutions différentes pourront s'imposer. Ce qui est aujourd'hui un outil parfaitement banalisé dans les services de la Commission commencera seulement à se généraliser dans d'autres environnements.

Une seule chose paraît sûre aujourd'hui : les développements technologiques et l'accroissement du parc informatique contribueront au développement et à l'élargissement de la gamme d'outils dont disposeront terminologues et traducteurs, mais le fond du problème, la vraie difficulté résidera toujours, par-delà l'outil informatique, dans le contenu de la base, dans le nombre et la qualité des données auxquelles voudront accéder les utilisateurs de terminologie, sans cesse plus nombreux et plus diversifiés.

20

La modernisation de la traduction à la Commission et ses moyens

Kees VAN DER HORST

Commission des Communautés européennes, Service de traduction, Luxembourg

Le cadre

Le Service de traduction

La Commission des Communautés européennes dispose d'un Service de traduction, organisé en sept unités thématiques, une Direction « Affaires générales et linguistiques » et quatre services horizontaux. Chaque unité thématique est composée de neuf unités linguistiques (danoise, allemande, hellénique, anglaise, espagnole, française, italienne, néerlandaise et portugaise). Chacune de ces 63 « unités de base » comprend de douze à trente traducteurs et un secrétariat. Au total, il y a quelque 1200 traducteurs et 500 secrétaires, dont un tiers environ est affecté à Luxembourg et deux tiers à Bruxelles. Ils produisent ensemble chaque année plus de 800 000 pages traduites.

Le défi européen

Aujourd'hui et demain, en ce qui concerne le Service de traduction, le défi n'est ni américain ni même japonais, mais bien européen. Les murs, les rideaux et les frontières tombent mais les langues et donc les barrières linguistiques en Europe resteront et deviendront plus que jamais visibles. La communication, cette vaste machine silencieuse et omniprésente, inaperçue par la plupart de nous comme l'eau l'est par les

poissons, avance à travers les anciennes frontières — à condition que la traduction suive...

Voici donc le défi européen pour le Service de traduction : s'assurer qu'aujourd'hui et demain la Commission des Communautés européennes trouve ses canaux de communication multilingues toujours ouverts.

Le concept clé : maîtrise

La traduction intervient à deux endroits précis dans la communication multilingue, soit à la source (avant transmission du message), soit à la destination. À la source elle ouvre, à la destination elle débloque autant de canaux de communication qu'il y a de traductions demandées. Dans les deux cas, la traduction subit les conséquences de tous les problèmes associés au message initial, y compris la multiplication de ces problèmes par le nombre de langues demandées en traduction. Malheureusement, la communication multilingue est rarement planifiée et organisée, et même si cela est le cas, les traducteurs n'y sont jamais associés.

Pourtant, la communication multilingue est un processus comme un autre. Elle a une structure et une dynamique que l'on peut examiner et décrire. Mais cela ne veut pas dire qu'elle est, en tant que processus, planifiée et organisée afin de se dérouler d'une façon efficace et contrôlée. Au contraire, comme presque tous les processus qui sont le résultat d'actions humaines, celui de la communication se déroule d'une façon relativement autonome et non planifiée. Le but que nous nous sommes fixé est de réduire cette autonomie relative et d'obtenir une certaine maîtrise sur son organisation et son déroulement.

En résumé, sans reconnaissance de l'impact des barrières linguistiques, il est difficile de faire apprécier le rôle de la traduction. Sans cela, il est difficile de conquérir la position sur la scène de la communication écrite qui permettrait à la traduction d'organiser et de maîtriser l'aspect multilingue. Et pourtant, c'est la seule voie qui conduise à une réduction progressive et systématique du problème de surcharge. Il faut remonter à la source, endroit où l'on peut intervenir avec le plus d'efficacité. Après transmission du message — le cas des traductions demandées à la destination — il est trop tard et il faut chercher d'autres moyens.

La stratégie

La stratégie proposée se compose de deux « offensives » :

La première est destinée à porter la traduction plus en amont dans le flux de la communication : obtenir que le Service de traduction soit reconnu et accepté comme un partenaire valable et indispensable dans la conception et la construction de voies de communication multilingues à la Commission. Dans cette optique, le Service de traduction ne jouera plus exclusivement le rôle passif de l'instance que l'on appelle au dernier moment pour « encoder » ou « décoder » des messages d'une langue à

d'autres. Il assume en parallèle le rôle actif de constructeur de systèmes de communication multilingues à la Commission.

Le but est clair : réduire, par un (long) travail préparatoire, l'intervention du Service de traduction dans la partie du besoin en traductions qui se prête à être ainsi traitée; notamment, une partie des traductions demandées à la source, avant transmission du message.

La deuxième « offensive » est le pendant de la première. Elle s'adresse à la partie du besoin en traductions que nous n'arriverons pas à maîtriser par la première offensive; en premier lieu, les traductions demandées par le destinataire du message.

Pour cette catégorie de textes, nous devons essayer de réduire, par la construction de systèmes propres aux traducteurs, la traduction du « déjà-traduit », la recherche de solutions aux problèmes de traduction déjà résolus.

Les moyens

Les outils sont importants, mais l'analyse précédente a démontré, je l'espère, que les mesures organisationnelles sont plus importantes que les moyens techniques. Renforcer la conscience de l'impact des barrières linguistiques, changer des habitudes et des attitudes et maîtriser la communication multilingue à travers une planification et une organisation efficace est un processus long et difficile.

En ce qui concerne la première offensive, nous sommes en train de développer des « aides à la rédaction de documents multilingues » (MLDDA — *Multilingual Document Drafting Aids*). Ce sont des outils informatiques dans lesquels nous incorporons du « savoir-faire » linguistique, afin de permettre aux auteurs de générer eux-mêmes les versions linguistiques de leurs documents ou de préparer leurs documents d'une façon qui allégerait très fortement la tâche de traduction. Nous cherchons également à identifier tous les circuits de communication multilingue à la Commission, dont les messages se prêtent à une transformation de texte libre en texte standardisé.

Pour la deuxième offensive, nous comptons développer des outils informatiques qui permettent aux traducteurs d'exploiter le patrimoine du « déjà-traduit ». Le premier pas est la création de systèmes centraux d'informations linguistiques, soit sous forme d'archives de traductions (stockage sous forme brute), soit sous forme de bases de données terminologiques, phraséologiques et documentaires (stockage sélectif). Ces systèmes (appelés le « *Translators' Memory* » — la mémoire des traducteurs) seront automatiquement exploités sur base du texte à traduire. Le résultat, sous forme de petites sous-bases terminologiques et documentaires ou du texte source avec des substitutions en langue cible, ou autre, sera transmis avec le texte à traduire par courrier électronique dans l'environnement personnel du traducteur (le « *Translator's Workbench* »). Cette architecture permet une très grande indépendance géographique — deuxième objectif important du Service de traduction, après la maîtrise. Déjà aujourd'hui, la moitié des traducteurs affectés à Luxembourg travaille pour des demandeurs de traductions à Bruxelles, en exploitant intensivement le courrier électronique, et cette dispersion géographique augmentera sans doute dans le futur.

Inévitablement, les tâches des traducteurs changeront à long terme. À côté de la traduction classique, qui restera dominante, les pionniers parmi les traducteurs participeront à la conception et au développement de systèmes de communication et d'information multilingues. Tous feront appel à des machines, afin de trouver dans cette symbiose la synergie indispensable pour relever le défi européen dans le domaine des langues : assurer la communication à travers les barrières linguistiques, quels que soient leur nombre et leur hauteur.

Pour ceux qui pensent que l'auteur rêve, voici une dernière information. Entre 1988 et 1991 les traducteurs de la Commission à Luxembourg ont conçu et mis en œuvre un système pour la maîtrise de la traduction des avis de marché publics, venant de tous les États membres et publiés chaque jour en neuf langues dans un supplément au Journal Officiel des Communautés européennes (volume : quelque 300 000 pages annuelles). En 1988, trois traducteurs sur quatre se consacraient à ce travail, tandis que la charge augmentait exponentiellement. Aujourd'hui, grâce à une standardisation par formulaires et nomenclatures, conçus et développés par les traducteurs, la participation du Service de traduction est réduite à quelques personnes. Une prochaine étape, basée sur le principe d'EDI (*Electronic Data Interchange*), pourrait éliminer toute intervention des traducteurs. Un système de communication, qui nécessitait plus de 150 traducteurs pour fonctionner, a été mieux planifié, mieux organisé et finalement transformé en un système qui travaille presque seul, avec une augmentation de la performance et de la qualité de la communication multilingue.

PARTIE III

Traduction, Terminologie, Informatique

Présidents : Raymond Renard
Daniel Godfrind
Guy Efoe Wallace
Benoît Ouoba

21

Aspects de l'informatisation des activités terminologiques et traductionnelles

Gabriel OTMAN

Centre de Terminologie et de Néologie, INaLF, CNRS, Paris, France

Au cours de la décennie écoulée la grande majorité des colloques consacrés à la terminologie ont traité de son informatisation : *L'ère nouvelle de la terminologie (Québec 1985), Terminologie et technologies nouvelles (Paris 1985), Séminaire de terminotique (Bruxelles 1986), Terminologie et ingénierie de la connaissance (Trèves 1987 & 1990), Terminologie et industries de la langue (Paris 1989), Applications micro-informatiques multifonctionnelles de la terminologie (Vienne 1989)* pour n'en citer qu'un échantillon représentatif.

Ces colloques sont doublés d'autant de halls de démonstration et de commercialisation de produits informatiques. C'est dire combien l'informatique liée au secteur d'activité de la terminologie et de la traduction suscite d'enthousiasme et d'intérêt qui s'expriment dans ce foisonnement de débats et de projets plus ou moins controversés. Nous ne nous rangeons pas d'emblée dans le camp des détracteurs de ces technologies nouvelles car nous ne réfutons pas l'affirmation que l'informatique peut être source de progrès, mais elle est aussi et d'abord source de fascination et de propos déraisonnables. Le fossé du possible à la réalité, du prototype à l'industriel est trop allégrement franchi par les marchands du temple. Nous proposons ici de verser à ce dossier quelques éléments de réflexion sereins exprimés avec une certaine distance à un moment charnière du développement informatique : le stade de la banalisation et de la massification des matériels précédant la période de leur incontournabilité.

Il est vrai que cette informatisation accélérée parfois frénétique que nous pouvons observer cherche à répondre à des attentes et des exigences des praticiens : exigences

de progrès techniques, exigences d'efficacité et de rentabilité et, en amont, exigences de formation.

Si l'informatique s'est banalisée sur un flanc, celui de la micro-informatique, elle s'est complexifiée sur un autre flanc, celui du traitement des connaissances. La banalisation a permis une multiplication et une diversification des utilisateurs par l'arrivée d'un public, de formation linguistique ou apparentée dans le domaine qui nous concerne, à peine initié à l'informatique. La crainte de ce public est de perdre la maîtrise d'un outil de travail dont la vitesse de croissance est exponentielle et ainsi d'avoir le sentiment intellectuellement insupportable de devenir un serviteur résigné de la technologie.

Si l'on considère que les performances intrinsèques des ordinateurs peuvent croître d'un facteur dix tous les cinq ans et que, parallèlement sur la même période, les marchés correspondants doublent en volume, la nécessité d'une réflexion en profondeur sur les effets et les enjeux de cette informatisation saute aux yeux. Ni pessimisme acharné ni optimisme forcené ne sont de mise car, quand bien même nos moyens d'infléchir la progression de l'informatisation sont limités sur l'ensemble du phénomène, ils sont réels dans la sphère qui nous concerne ne serait-ce que par le choix entre adoption ou rejet de certaines options matérielles et logicielles.

Observons donc quelques-uns des aspects de l'informatisation sur lesquels doit particulièrement porter notre attention car, même si notre connaissance de la science informatique reste superficielle, nous pouvons influer sur ses manifestations extériorisées telles que l'ergonomie logicielle ou la formation des usagers non spécialistes.

Si l'informatisation est inéluctable, elle est néanmoins multiforme et peut prendre l'aspect qui, ergonomiquement et productivement, s'apparente au mieux aux attentes de chacun. En terminologie, dans la mesure où l'option du « grand chaudron », du système centralisé hyperpuissant et hypercomplexe qui prévalait au début de la dernière décennie est aujourd'hui abandonnée, à preuve la multiplication des mini-banques de données terminologiques locales face aux mastodontes interrogeables à distance, le développement des « petits chaudrons » individualise le processus d'informatisation et responsabilise chaque utilisateur.

L'informatique en changeant d'échelle a changé de nature. Elle est capable de traiter des problèmes d'un type nouveau : on est ainsi passé des nombres aux textes, du calcul mathématique au traitement de chaînes de caractères, de la déduction algorithmique au raisonnement incertain sur des données incomplètes, voire à l'apprentissage automatique. Nous sommes au cœur d'une phase d'optimisation. Si la croissance se fait sur l'ensemble, l'optimisation, quant à elle, se fait au coup par coup sur les parties. Cette orientation vers l'optimisation passe par la prise de conscience de la faillibilité de la machine (pannes, erreurs, etc.), de la probabilité non nulle que les résultats que la machine fournit puissent être entachés d'un facteur d'erreur.

Par ailleurs, le lien entre le chercheur, la langue et l'informatique a été institutionnalisé par la création de filières du type « informatique linguistique » ou « linguistique computationnelle ». L'informatique, avec l'intelligence artificielle, est passée pour le linguiste de levier à sujet de recherche. Ce changement d'orientation a provoqué

le recours nécessaire à des modèles formels et des traitements extrêmement contraints par la machine, ce dont certains linguistes à l'esprit rigoureux se félicitent.

Mais avant de considérer son matériel sous l'angle potentiel d'outil d'investigation théorique, l'« informatisé » acquiert un outil de travail qu'il veut à la fois, comme l'annonce la publicité, « performant et convivial », et intellectuellement gratifiant.

Informatisation et qualité de travail

L'informatisation induit un nouveau mode d'organisation du travail qui passe par la recherche de trois types de tâches : les tâches entièrement automatisables, les tâches partiellement automatisables où la machine assiste l'humain et les tâches nobles, créatives, strictement réservées à l'humain. En effet, si la traduction reste un art (au même titre que la terminologie selon Louis-Jean Rousseau) et si la rédaction de fiches terminologiques laisse encore une part à la créativité, ces deux activités comportent une quantité appréciable de tâches mécaniques et routinières que la profession gagnerait à automatiser. Ainsi, à l'Université Laval, une équipe de chercheurs, autour de Pierre Auger, s'est penchée sur l'automatisation de l'ensemble de la chaîne de travail terminographique, chaîne comprenant les tâches de sélection de documents, de dépouillement et de repérage de termes, d'établissement de nomenclatures et de sélection et traitement de contextes. Plus généralement, en bureautique, on estime à 40 pour cent le taux des activités informatisables. Norbert Kalfon (1989 : 610) estime à 70 pour cent la part du travail automatisable en traduction : recherche terminologique, préparation de textes, correction orthographique et stylistique, post-édition, etc.

Cette nouvelle distribution entraînera à moyen terme une baisse de qualification pour le plus grand nombre et une surqualification pour le groupe placé au sommet de la pyramide, celui qui maîtrise les développements technologiques et stimule l'informatisation. De la massification des implantations informatiques résulte une exigence sociale à participer à cette informatisation, de bon gré pour les uns qui s'y adaptent aisément et y reconnaissent un élément de progrès, de mauvais gré pour les autres qui y résistent et vont grossir les rangs des « technopathes », ces personnes réfractaires à l'informatique en général et à l'idée de communiquer interactivement avec un ordinateur en particulier.

Il y a fort à parier que les terminologues-traducteurs atteints de technopathie, s'il en existe encore, seront condamnés à disparaître à très court terme. Des enquêtes sérieuses ont prouvé que les entreprises informatisées voyaient l'âge moyen de leur personnel diminuer. L'informatisation à outrance peut donc éliminer des catégories sociales où le taux de technopathes est plus élevé : les personnels âgés et féminins en particulier. Ce qui, à terme, ne peut être que préjudiciable à une discipline touchant aux sciences humaines où l'expérience acquise et l'intuition linguistique priment sur la performance technique.

Associer systématiquement la notion d'informatisation à celle d'amélioration des conditions de travail et de la qualité de la production, c'est aller un peu vite en besogne. En effet, avec l'informatisation est apparue une nouvelle forme de pénibilité pro-

fessionnelle : pénibilité physique (maux de tête, acuité visuelle, mal de dos...), pénibilité psychique et mentale (le « stress informatique »). En effet, l'informatisation se traduit par des cadences de travail plus élevées, imposées par la hiérarchie ou, le plus souvent, imposées volontairement. Si l'on considère, par exemple, qu'un traducteur peut traduire une moyenne de 250 mots/heure, il peut réviser 1000 à 1500 mots prétraduits par une machine dans le même laps de temps. À combien de fiches le terminographe bien équipé sera-t-il astreint? Sur un plan strictement économique, ces gains de productivité sont partiellement ou totalement annulés par une dévalorisation du produit et des personnes qui en sont les producteurs.

Oui, mais ces inconvénients — mal mesurés encore — ne sont-ils pas contrebalancés par les possibilités de déconcentration et d'extériorisation des enceintes de travail qu'autorise l'informatique, le trop fameux « télé-travail »? Il semblerait au contraire que cette atomisation des lieux de travail, qui déboucherait sur un isolement pénible, ne soit souhaitée ni par les décideurs ni par les acteurs.

Par contre une demande s'exprime unanimement pour des produits plus ergonomiques, l'ergonomie minimale étant la compatibilité entre matériels, entre matériels et logiciels et entre logiciels. Des progrès sont réalisés dans ce domaine par la réalisation d'interfaces entre traitements de texte et dictionnaires, entre traitements de texte et courrier électronique par exemple. L'ergonomie (du grec « ergon » travail) est définie comme la science de l'homme au travail, l'ergonomie informatique étant dès lors l'étude des conditions de travail de l'homme confronté à l'outil informatique et, par voie de cause à effet, la recherche de l'amélioration de ces conditions de travail ou, à tout le moins comme cela semble être le cas actuellement, leur non-détérioration.

L'ergonomie informatique ne peut plus être traitée convenablement par l'utilisateur au stade du choix des matériels mis à sa disposition sur le marché. Même si le souci ergonomique reste un paramètre important au moment de l'achat, ce ne peut plus qu'être alors l'option du pis-aller de la non-détérioration. L'ergonomie est bel et bien un paramètre de conception au même titre que la portabilité ou la compatibilité. L'ergonomie matérielle a atteint certaines limites et un niveau acceptable de satisfaction (systèmes d'exploitation, miniaturisation, extension de mémoire, normalisation des configurations, terminaux multifonctions, etc.), c'est l'ergonomie logicielle qui pose problème. À mon sens, au niveau de sophistication auquel nous sommes parvenus, toute amélioration de l'ergonomie logicielle doit obligatoirement passer par la définition d'un certain nombre de normes auxquelles devront obéir tous les produits destinés au même usage. Par exemple, l'interrogation de banques de données terminologiques quelles qu'elles soient devrait être réalisable par le truchement d'un langage quasi naturel et par des fonctions semi-transparentes non contraintes par les concepteurs de ladite banque aussi bien pour la consultation que pour l'échange de données. Cet effort, s'il est généralisé, pourra conférer à une gamme entière de logiciels un véritable air de famille et un sentiment agréable de familiarité et de convivialité pour ses utilisateurs. Toutefois, ces éléments de normalisation ne doivent en aucun cas être imposés par le plus fort au plus faible selon les lois classiques du marché. Il serait préférable qu'elles résultent d'une étude ergonomique menée scientifiquement en collaboration avec les

usagers pour exploiter le plus économiquement possible les capacités techniques de l'ordinateur et les performances intellectuelles de l'humain.

Or, ergonomie et économie ne font pas bon ménage.

Informatisation et qualité de production

Terminographie et traduction sont des activités à rendement obligé et aucun traducteur indépendant ne nous contredira sur ce point. Le rendement s'exprime par une productivité accrue qui peut se comptabiliser en fiches réalisées ou en pages traduites. Notre secteur d'activité n'appartenant pas au secteur industriel purement productif mais plutôt au secteur des services, le rendement doit aussi s'exprimer par un accroissement qualitatif. Ce progrès qualitatif peut provenir de l'aide fournie par l'ordinateur puisque ce dernier est en mesure de soumettre au traducteur-terminologue un ensemble d'éléments d'information suffisamment conséquent et pertinent pour qu'il puisse prendre la meilleure décision linguistique possible en un minimum de temps.

Ce qui revient à dire qu'un investissement informatique, comme tout investissement d'infrastructure, a un coût qui, d'une part se répercute directement sur le produit et qui, d'autre part, doit entraîner un gain de productivité tel que le coût du produit fini n'augmente pas mais baisse relativement. Les spécialistes nomment « retour sur investissement » ce principe de base de la justification d'investissement. À ma connaissance, il n'existe pas de chiffres ni de moyens de mesure des gains de productivité en terminologie ou en traduction. En bureautique par contre, une étude du Crédit Agricole situe à 30 pour cent de gain de productivité le seuil de rentabilité d'une station de travail de 45 000 F. Sans vouloir nous lancer dans une bataille de chiffres faussée d'avance du fait qu'aucune méthode de mesure de gain de productivité reconnue n'est admise par tous, souvenons-nous simplement que cette dimension devra être de plus en plus prise en compte au moment de s'équiper.

On nous parle souvent de baisse de moitié des prix du matériel tous les trois ou quatre ans. Si cette baisse brute est irréfutable, elle n'est malheureusement que partiellement répercutée au niveau des investissements dans la mesure où, heureusement cette fois, à coût égal, on peut — et on veut — investir dans des matériels plus puissants et des logiciels plus performants. Il en résulte que la masse financière nette d'investissement ne marque pas le pas mais au contraire ne cesse de progresser, l'offre étant de plus en plus attirante.

Outre le coût du matériel, notre secteur d'activité est particulièrement sensible au coût de l'information produite. Ce dernier peut très rapidement dépasser le coût du système qui la supporte, sans même oser mentionner la perte irréparable et inestimable que peut représenter la destruction accidentelle de cette information. La sécurité, heureusement de moins en moins négligée, a également un coût qui peut être exorbitant en cas de duplication complète d'un service informatique. On se contente généralement d'un compromis économiquement viable, techniquement réalisable et intellectuellement satisfaisant s'il représente un juste équilibre entre les risques et les enjeux.

Pour compléter le tableau financier, n'omettons pas de mentionner qu'une bonne part de l'investissement informatique peut être gaspillée par une sous-utilisation chronique des matériels qui peut aller jusqu'à 30 pour cent de ses capacités potentielles. Cette marge variable entre capacités techniques et utilisation pratique n'est jamais négligeable et est trop souvent si conséquente qu'elle ronge la meilleure part de la marge de productivité.

Comment y remédier? Par une meilleure formation initiale et continue prenant réellement en compte la dimension d'informatisation croissante de nos activités car l'épanouissement d'une discipline, d'un secteur d'activité et de recherche dépend directement de la qualité de la formation donnée et de ses facultés d'adaptation.

Informatisation et qualité de formation

Certains services de terminologie et de traduction ne sont pas informatisés ou sont sous-informatisés non pas en raison d'une prise de position résultant d'une option déontologique sur la question de l'informatisation mais faute de pouvoir recruter un personnel qualifié à la fois sur le plan linguistique et sur le plan informatique. La question de la formation initiale est donc à la fois cruciale et urgente car c'est souvent le seul sentiment de dépossession du savoir qui provoque le rejet d'une technologie.

L'appareil de formation, principalement universitaire, doit remplir un double rôle : dans un premier temps une alphabétisation informatique couplée à un enseignement théorique et pratique des techniques et des méthodologies terminographiques et traductionnelles et, dans un second temps, un recyclage régulier des personnels formés en activité. Cette double formation doit allier connaissances théoriques et compétences pratiques. Si l'on se réfère au panorama très complet et précis que nous a offert la seconde livraison de *Terminologies Nouvelles* (décembre 1989) entièrement consacrée au thème de la formation en terminologie, il faut se rendre à l'évidence qu'à de rares exceptions près nous n'allons guère au-delà de l'alphabétisation et de l'information minimale concernant les matériels et les ressources disponibles : modalités d'interrogation des banques de données terminologiques, dictionnaires électroniques, SGBD et SGBDR, outils d'aide à la compilation de lexiques, interfaces, etc.

Pourtant, si l'on examine les attentes formulées par les stagiaires en formation continue, c'est à la terminographie assistée par ordinateur qu'ils veulent être formés en priorité. La formation doit continuellement s'adapter au contexte dans lequel les compétences acquises sont investies et aux évolutions de ce contexte. De plus, le terminologue-traducteur souhaite pouvoir s'engager de plus en plus dans la conception et le développement des outils informatisés qui lui sont destinés. Sa formation doit lui fournir les connaissances nécessaires à un dialogue possible avec des ingénieurs logiciels. Ce dialogue permettrait notamment aux concepteurs de systèmes de TA ou de TAO d'explorer certains mécanismes cognitifs de la traduction humaine. Il doit aussi pouvoir pénétrer les programmes, les comprendre suffisamment pour les maintenir au sens informatique du terme, les maîtriser et les faire évoluer. Par exemple, le traducteur face à son système de TAO ne peut qu'accéder à la base de données lexicales et

terminologiques qu'il peut alimenter à sa guise tandis que les modules comportant les règles syntaxiques et grammaticales ne lui sont pas accessibles même s'il y voit des imperfections auxquelles il peut apporter des améliorations. Les langages de programmation qualifiés de déclaratifs (au contraire des langages dits procéduraux) parce qu'ils permettent d'écrire des règles de production indépendantes les unes des autres — sous une forme déclarative — sont un moyen à la portée du traducteur formé et averti pour tester sans grands risques ni frais ses intuitions linguistiques. Il suffirait que les concepteurs ouvrent une voie d'accès à ce type d'intervention dans leurs programmes.

La lenteur d'adaptation de l'appareil de formation est sa principale tare face à la vitesse d'évolution du monde informatique. De plus, c'est uniquement par le levier de la formation que pourront être résolus certains aspects de l'harmonisation des méthodologies que souhaitent la plupart des praticiens de la terminologie géographiquement éparpillés dans les pays francophones. La formation se doit également d'être une porte ouverte vers d'autres disciplines connexes telles la psychologie, l'intelligence artificielle ou les sciences cognitives dont le contact et le frottement sont le ferment de la recherche fondamentale.

Et qu'en sera-t-il demain quand les compétences du terminologue-traducteur s'étendront à tout ce que nous prépare la terminotique et la traductique?

En effet, en nous autorisant une formule quelque peu osée, nous pouvons dire que, jusqu'aux années quatre-vingt-dix, l'informatisation faisait la part belle aux comportements *réflexes* et, qu'à partir de la décennie qui s'ouvre, elle s'attaque de plain pied aux comportements de *réflexion*.

Regards prospectifs sur l'informatisation de demain

Les axes de développement principaux des recherches actuelles nous conduisent vers une formalisation et une systématisation plus poussées du travail terminologique et traductionnel qui correspondent aux quatre activités principales du terminologue-traducteur : la recherche, la consultation, le tri et la transcription.

Avec l'émergence des applications de l'intelligence artificielle dans notre secteur d'activité, nous passons de l'ère du traitement de l'information à celle du traitement des connaissances; le saut est, ici aussi, autant qualitatif que quantitatif puisqu'on n'ambitionne pas moins que d'emmagasiner la « connaissance du monde » et de la rendre accessible sous les formes les plus diverses aux utilisateurs les plus nombreux possible.

Concrètement et très progressivement cela se manifestera par un accès facilité à un grand nombre de dictionnaires ou de bases de données à partir d'un simple terminal dans un mode d'interrogation homme-machine se rapprochant de plus en plus du langage naturel, par la constitution de bases de données relationnelles évoluées prenant de plus en plus en compte les relations sémantiques entre notions par le recours à des structures fondées sur les réseaux sémantiques par exemple, ou encore par le développement de la gamme des logiciels d'hypertextes qui ouvriront la voie à une vraie recherche multicritère.

Ces nouvelles bases de connaissances seront dynamiques, c'est-à-dire en croissance permanente par l'apport de connaissances grâce à des fonctions d'auto-apprentissage. Cet auto-apprentissage se fera entre autres par la ponction automatique de connaissances contenues dans des textes à partir de systèmes d'analyse et de compréhension du langage naturel regroupés en progiciels intégrés.

Le poste de travail du terminologue-traducteur évoluera en conséquence. On y aura incorporé des analyseurs syntaxiques et grammaticaux qui viendront épauler des correcteurs stylistiques. Il sera multitâche et permettra de jongler entre une multiplicité de sources d'information d'une part et d'outils de génération de textes d'autre part afin de réduire au minimum la durée entre entrée d'information et sortie du produit final sous la forme la plus élaborée possible. Le problème ne sera plus l'absence d'information ou la difficulté d'accéder instantanément à l'information requise mais la gestion du flux d'information. Il faudra éviter les effets de blocage dus à un engorgement du système par trop-plein d'information. Ici, la réflexion humaine génératrice de critères de tris affinés et pertinents sera capitale pour l'épuration de ce flux. Ces bouleversements de l'environnement de travail, des fonctions et des attributions entraînera une nouvelle définition des acteurs que nous sommes; l'appellation est déjà prête pour les uns, terminoticiens. On peut risquer « traducticien » (de traductique) pour les autres.

Ces systèmes nourris des résultats des recherches en intelligence artificielle sur le raisonnement et le fonctionnement cognitif humains seront en mesure de traiter la connaissance en simulant nos propres heuristiques mentales de spécialistes par des programmes contenus dans des biopuces.

Mais voilà que nous aussi, à l'image des gourous de l'informatisation, nous nous laissons entraîner vers les sphères nébuleuses d'une prospective lointaine et incertaine dans laquelle l'ordinateur égale ou dépasse les facultés humaines. Arrêtons-nous là sous peine de nous faire taxer de « douce rêverie ».

Conclusion

Comme tous les progrès, l'informatique mêle espoirs et menaces. Il n'en demeure pas moins qu'aujourd'hui, quoi qu'on puisse penser des dangers de l'informatisation, on ne cesse d'exiger toujours davantage de ces étranges machines, toujours plus de performances au-delà des limites de ce qu'il peut sembler raisonnable d'attendre d'un appareil qui est et restera une machine inerte.

L'emprise de l'informatique ne doit pas empiéter sur l'action de l'homme, ses modes de pensée et sa faculté de créativité. Mais ce n'est qu'en hissant l'informatique à lui et non en se baissant vers elle que l'homme assurera sa maîtrise sur cette technologie. Et alors peut-être une machine à qui on demandera — en langage naturel — une collocation de « informatisation » et « aliénation » répondra que ces deux termes ne sont co-occurrents dans aucune phrase contenue dans un document daté de 2000 ou au-delà, alors que le terme « informatisation humaine » s'est banalisé.

Bibliographie

TERMINOLOGIES NOUVELLES (1989) « La formation en terminologie », Bruxelles, ACCT, Communauté française de Belgique, décembre 1989, n° 2.

META (1989) Actes du colloque « Terminologie et Industries de la Langue », vol. 34 n° 3, septembre 1989.

META (1987) « Vers l'an 2000, la terminotique : bilan et prospectives », vol. 32 n° 2, juin 1987.

DE SCHAETZEN Caroline (1990) « Outils de bureautique et de télématique pour la traduction », *Lebende Sprachen* n° 3/90, p. 123-128.

22

Comparaison de logiciels utilisables en terminologie

Élisabeth BLANCHON

Centre de Terminologie et de Néologie, INaLF, CNRS, Paris, France

Introduction

À une époque où, c'est un truisme, l'informatique, censée résoudre tous les problèmes, a envahi tous les domaines de la connaissance et se rencontre dans tous les environnements de travail, on se trouve confronté à un nouveau problème, celui de la multiplicité des offres logicielles : quel logiciel choisir, et sur quels critères?

On répondra d'abord par ce qui devrait être une évidence : il faut avant tout procéder à une analyse de ses besoins (et de ses moyens, mais ce deuxième aspect s'impose de lui-même). En l'occurrence, on adoptera l'optique du traducteur qui dispose déjà d'un traitement de texte, sur micro-ordinateur compatible PC, et qui souhaite consulter ses données terminologiques sans quitter son logiciel, ou le moins possible, et insérer dans son traitement de texte des données de ses fiches terminologiques. En outre, ce traducteur lambda ne dispose pas nécessairement de l'aide permanente d'un informaticien ni ne souhaite apprendre lui-même à programmer : il a besoin d'un logiciel suffisamment convivial pour être directement utilisable ou ne pas exiger un apprentissage trop long.

Pour commencer on citera les principaux types de logiciels effectivement disponibles sur le marché, et utilisés en terminologie, avec pour chacun, ses avantages et ses inconvénients.

Dans un deuxième temps, on examinera un peu plus en détail les spécificités de quelques-uns de ces logiciels.

Enfin, on conclura par une sorte de cahier des charges du logiciel idéal, qui est une sorte de mise en perspective des besoins réels du traducteurs et des possibilités réelles de ces logiciels.

Les principaux types de logiciels utilisables pour gérer des données terminologiques

En réalité, n'importe quel logiciel de gestion de données, ou presque, peut être détourné de sa vocation d'origine, si tant est qu'il en ait eu une, pour être utilisé en terminologie. En effet, même si elles présentent certaines particularités, les données terminologiques ne sont que des données, et une fiche terminologique peut être traitée de la même manière qu'une fiche produit ou une fiche client.

On a donc le choix, comme dans tout autre domaine, entre les logiciels traditionnels, traitement de texte et systèmes de gestion de bases de données, et des logiciels spécifiques, en l'occurrence terminologiques.

Les traitements de texte

On peut se limiter à deux grand noms : Word 5 et WordPerfect 5, tout traitement de texte digne de ce nom permettant de réaliser les mêmes opérations, sous des dénominations variées. Il faut, en effet, pour être utilisable en terminologie, qu'un logiciel de traitement de texte permette d'enregistrer et de rappeler une structure de fiche vide, de rechercher des mots, d'afficher des fenêtres à l'écran, et offre une fonction couper-coller, toutes fonctions assez largement partagées par les traitements de texte courants.

Si l'on admet l'intérêt, largement accepté, de réaliser une FICHE terminologique, à structure bien établie, on peut préparer un document spécifique, une structure vide, reproductible à volonté, que l'on remplira au fur et à mesure des besoins, comme une fiche normale. Sous Word, on utilisera la fonction GLOSSAIRE, au moyen des commandes *Recopie* ou *Détruit*, en donnant un nom, par exemple FICHE, à cette fiche vide. Quand on voudra l'utiliser, il suffira de taper FICHE et la touche de fonction F3 pour faire apparaître cette structure vide. On peut sauvegarder ce masque comme n'importe quel document. La fonction équivalente dans WordPerfect est la fonction MEMO.

Le traitement de texte laisse également la liberté de réaliser des fiches sans structure uniforme, ce qui peut parfois être souhaitable.

Dans un cas comme dans l'autre la recherche se fera au moyen de la commande de recherche (*Cherche*, dans Word).

Le fichier terminologique n'est alors rien d'autre qu'un document parmi d'autres, traité par le logiciel comme n'importe quels lettre ou article.

Le traducteur peut passer de son document de traduction à son fichier terminologique en créant deux fenêtres (au moins) à l'écran, en allant de l'une à l'autre en

pressant une touche (F1 sous Word) pour faire une recherche, découper les informations souhaitées, et les recoller dans le document de traduction.

L'INTÉRÊT d'utiliser un traitement de texte pour gérer ses données terminologiques est évident :

- on n'a pas besoin d'acquérir un autre logiciel;
- on échappe aux problèmes d'interfaçage, de transfert de données, etc.;
- on est libre d'élaborer la fiche que l'on souhaite et de l'utiliser ou non.

C'est donc un choix de moindre coût et de simplicité et de liberté accrues.

En outre, si l'on désire ultérieurement, devant l'abondance de sa production terminologique, publier ses données sous forme de dictionnaire, on pourra directement les mettre en forme dans ce but, sans étape supplémentaire.

Les INCONVÉNIENTS sont d'un tout autre ordre :

- la recherche de termes, dans un traitement de texte, ne peut se faire que de manière séquentielle (par chaîne de caractère), ce qui peut être assez long si on le fait sur un document important;
- la recherche ne pourra se faire que sur ce seul critère d'identité de chaîne de caractère : on ne pourra pas extraire, par exemple, l'ensemble des fiches d'un même domaine, ou réalisées par un même auteur ou depuis telle date, sauf en réitérant la recherche et en recopiant dans un autre document les fiches ainsi récupérées.

En réponse aux deux inconvénients cités ci-dessus on peut toujours envisager d'avoir plusieurs « bases de données », une par domaine, par exemple.

Le problème majeur est un problème de recherche : on y perd en vitesse et en précision.

On peut citer l'Université de Rennes et le Conseil international de la langue française (CILF), entre autres, comme utilisateurs de traitements de texte pour gérer des bases de données terminologiques.

Les systèmes de gestion de base de données (SGBD)

Il nous paraît souhaitable de distinguer les SGBD généralistes (ou non spécifiques), qu'ils soient classiques ou relationnels, et les SGBD documentaires.

Généralistes ou spécifiques, les SGBD ne permettent généralement pas un transfert aisé des données vers un traitement de texte.

Les SGBD généralistes

Parmi les SGBD classiques, nous pouvons citer Reflex de Borland, dBaseIII + d'Ashton-Tate, Foxbase + de Fox Software. Parmi les relationnels, Paradox de Borland, dBase IV d'Ashton-Tate, Oracle de la société Oracle.

Leur inconvénient majeur est la faible longueur de leurs champs (254 caractères maximum pour les champs interrogeables de dBase ou Foxbase), généralement trop brefs pour gérer des informations textuelles comme la définition ou le contexte.

L'intérêt majeur de certains d'entre eux est d'être extrêmement répandus, au point d'être devenus des standards de fait, ce qui facilite les échanges et les conversions de données, car presque tous les logiciels ont prévu d'intégrer le format dBase, pour ne pas le citer, dans leurs formats d'importation/exportation.

Ils présentent également l'avantage de permettre des tris multicritères qui offrent la possibilité de sélectionner des sous-ensembles de la base.

Les SGBD documentaires

Il peut paraître singulier de mentionner ce type de logiciels, mais rappelons que la BTQ et les bases terminologiques d'IBM tournent sur Stairs, logiciel documentaire d'IBM. En outre, leurs caractéristiques sont assez différentes de celles des SGBD non spécifiques.

On citera surtout les logiciels TEXTO (Chemdata, Lyon), CDS-Isis de l'Unesco, et EDIBASE d'Inform II Microfor.

Leur gros avantage est d'accepter des champs généralement beaucoup plus longs que les SGBD généraux et de permettre une recherche multicritère par opérateurs booléens ET, OU, NON, plus simple, avec généralement un historique des questions et la possibilité de reprendre des étapes antérieures pour les combiner avec d'autres questions.

Autres logiciels

Il faut encore dire un mot des logiciels intégrés, comme Framework, Symphony ou encore Works. En effet, leur SGBD et leur traitement de texte étant parfaitement compatibles, ils permettent d'éviter bon nombre de difficultés de transfert, au prix néanmoins d'une moindre qualité du traitement des données textuelles.

On peut également, comme on le disait au début, aller jusqu'à détourner de sa vocation d'origine un tableur comme Excel ou Lotus, pourquoi pas, avec bien sûr les limitations inhérentes à ce genre de logiciel (taille des champs, etc.).

Enfin on ne peut pas passer sous silence un Deskview ou un Windows, qui permettent la consultation en parallèle (par multifenêtrage) de plusieurs applications, sans toutefois permettre de transfert facile de données, sauf, dans le cas de Windows, entre des applications qui sont prévues pour tourner sous Windows.

Les logiciels terminologiques

Bien qu'on ait choisi de les distinguer des précédents, il faut bien rappeler que ce sont aussi des logiciels de gestion de données. Ils sont censés être adaptés au traitement

des données terminologiques, mais on se rend compte que ce traitement donne lieu aux interprétations et aux réalisations les plus diverses.

On peut en effet distinguer quatre catégories essentielles, combinables entre elles : les logiciels qui permettent un accès direct à partir d'un logiciel de traitement de texte, ceux qui n'offrent pas cette possibilité, ceux qui imposent un format de fiche prédéfini, ceux qui autorisent l'utilisateur à définir son propre format de fiche.

	Accès direct à partir du traitement de texte	Pas d'accès direct à partir du traitement de texte
Fiche libre	Termex Multiterm	LexM
Fiche prédéfinie	Term-tracer Superlex Profilex Term-PC CATS	Aquila MC4 Index

Autant dire que les avantages et inconvénients respectifs de ces logiciels dépendent du type auquel ils appartiennent. Il n'y a en effet aucune commune mesure, en matière d'ergonomie, entre un logiciel à fiche libre et pouvant être consulté directement à partir d'un traitement de texte et un logiciel à fiche prédéfinie, dont les champs, plus ou moins nombreux, correspondent plus ou moins bien aux besoins du traducteur, et qui ne permet pas la consultation à partir d'un traitement de texte.

Quelques logiciels effectivement utilisés en terminologie

On laissera de côté les logiciels de traitement de texte, dont on a dit quelques mots plus haut, mais dont certains sont bel et bien utilisés en terminologie.

D'autre part, je tiens à me limiter aux logiciels avec lesquels j'ai eu personnellement l'occasion de travailler, dans les proportions les plus diverses il faut bien le dire, c'est-à-dire Aquila, MC4, et Termex pour les logiciels de gestion de données terminologiques, Foxbase + comme SGBD non spécifique et Edibase, Isis et Texto comme logiciels documentaires.

Parmi ces logiciels, seul Termex autorise une consultation directe à partir d'un traitement de texte.

D'autre part deux des logiciels terminologiques, MC4 et Aquila, imposent une fiche prédéfinie, sans permettre l'adjonction d'aucun autre champ.

Dans tous les cas, on peut effectuer des recherches sur les termes, sur les listes de termes (index) et des recherches par chaînes de caractères. Il convient toutefois de

noter que les différents logiciels gèrent de manière variée les index ainsi que la recherche par chaîne de caractères. Ainsi Aquila, MC4 et Termex ne proposent d'index que sur les vedettes, alors que dans les autres, il est possible de créer un index sur chacun des champs, et même, dans le cas des trois logiciels documentaires, sur des mots présents à l'intérieur des champs. Quant à la recherche par chaîne de caractères, Foxbase ne l'accepte que champ par champ, alors que Termex, Isis et Edibase peuvent la faire sur l'intégralité de la fiche.

Dans tous les cas aussi, on pourra, ou l'on devra, envisager le transfert des données vers un traitement de texte pour aboutir à un résultat plus adapté aux exigences éditoriales.

Spécificité de ces logiciels

Aquila et MC4

Le plus gros reproche que l'on puisse faire à ces deux logiciels, outre l'absence de consultation possible à partir d'un traitement de texte, est de contraindre l'utilisateur à se conformer à une structure de fiche très stricte et très limitée. Les champs prévus ne correspondent pas nécessairement aux besoins du traducteur, et ne sont pas assez longs pour que l'on puisse en quelque sorte tricher.

Aquila répond à un souci terminologique d'entreprise, et de ce fait nomme contexte ce qui est un ensemble d'informations sur l'entreprise, la filiale et le produit dans lesquels on trouve le terme vedette.

MC4 présente l'intérêt louable de correspondre au format de la banque de données terminologiques EURODICAUTOM et de permettre donc une importation directe dans EURODICAUTOM. Il permet théoriquement une gestion hiérarchisée des notions, sous forme de réseau; en réalité, l'expérience prouve que cette possibilité rencontre assez vite ses limites. D'autre part, on peut lui reprocher une certaine lourdeur dans la gestion des différentes fonctions, en particulier l'absence d'accès à l'intégralité de la fiche, lourdeur liée à un souci extrême de sécurité. En effet, les différentes fonctions, de saisie, de modification, d'édition ou impression, relèvent de « modules » différents. Un mot de passe est même prévu pour pouvoir accéder au module de saisie et modification, et le logiciel demande confirmation sur confirmation et information sur information, ce qui devient fastidieux dans le cas d'un travail terminologique suivi.

Parmi les utilisateurs de MC4 on peut noter le Service de terminologie du ministère de l'Économie et des Finances français.

Nous n'avons pas d'information sur les utilisateurs d'Aquila.

Termex

C'est le seul de tous les logiciels examinés à permettre à la fois un accès direct à partir d'un traitement de texte (Word, WordPerfect, Wordstar 2000 et bien d'autres) et une structure de fiche libre.

Au risque de paraître d'un enthousiasme suspect, il faut admettre qu'il est de ce fait étonnamment souple d'utilisation : on peut à tout moment définir ou réactiver un autre masque de saisie, ou choisir de travailler sans masque. La structure de la fiche est en effet considérée elle-même comme une fiche.

On peut créer des liens de fiche à fiche, l'un de ces liens pouvant être rendu automatique par un menu d'options, les autres se constituant à la demande, du fait de l'identité du contenu d'un champ avec la vedette d'une autre fiche.

La possibilité d'appel à partir d'un traitement de texte est accompagnée d'une possibilité de découpage d'éléments de fiches pour les réinsérer dans le texte, ainsi que de la possibilité d'utiliser le mot sur lequel on se trouve dans le texte pour faire une recherche directe sur lui ou le traiter comme nouvelle vedette.

Deux fichiers peuvent être ouverts en parallèle : un fichier de travail, que l'on peut modifier, et un fichier d'archivage, protégé, un fichier de travail « parfait » pouvant être transformé en fichier d'archivage. Des bases de données terminologiques peuvent également être acquises pour certains domaines. Signalons également la possibilité de conversion vers le format en cours de normalisation internationale Micromater, ainsi qu'à partir de celui-ci.

Dans le cas d'un fonctionnement en réseau, des niveaux d'accès sont prévus pour gérer la base et protéger son intégrité.

Enfin, d'autres outils encore font de ce logiciel un outil très intéressant dans le cadre d'une activité de traduction.

On pourrait reprocher à Termex de ne pas permettre de tris multicritères, mais il semblerait que cette fonctionnalité soit à l'étude. Enfin, comme on devient exigeant dans un cas pareil, on pourrait également regretter que des possibilités d'édition réduites nécessitent un retraitement sur traitement de texte pour publier des dictionnaires.

Utilisateurs : entre autres, le Secrétariat d'État du Canada, la Cour de justice des Communautés européennes, le ministère des Postes français.

Foxbase +

Ce que l'on peut dire ici de Foxbase +, qui, il faut bien le dire, est un cousin germain ou un frère jumeau de dBase III +, exclut tout souci de programmation. Il est évident que l'utilisateur qui est prêt à consacrer un certain temps à programmer ce logiciel en fonction de ses besoins peut en tirer à peu près n'importe quoi, en particulier s'il adopte la toute dernière version Foxbase Pro.

Nous lui reprocherons avant tout la brièveté de ses champs consultables (254 caractères). Il existe, nous opposera-t-on, des champs Memo, acceptant jusqu'à 5000 caractères, mais il est impossible de procéder à une recherche sur ce type de champ.

Il autorise la plus grande liberté de conception de fiches, et même de modifications ultérieures. D'aucuns pourraient lui reprocher, surtout par rapport à MC4, l'absence de sécurité et le risque de perte d'information, que ce soit par réécriture involontaire ou par restructuration trop rapide de la fiche. En effet saisie, consultation et modification

ne sont qu'un seul et même module et l'on pourrait assez facilement détériorer des données valides par une mauvaise manipulation à la consultation, et la possibilité de redéfinition de la structure de fiche peut, si celle-ci n'est pas réalisée dans les normes, parfois implicites, engendrer des disparitions de données. En réalité, tant qu'il ne s'agit que du travail d'une seule personne, et pas de celui d'une équipe, il peut même être préférable, pour une question de simplicité de procédures, et de rapidité de saisie, de ne pas avoir trop de contraintes de sécurité.

Utilisateurs : Centre de terminologie et de néologie (CTN), CERTIL.

Texto

Plus couramment utilisé en documentation, sa définition d'origine, ce logiciel peut être très facilement détourné de sa vocation première, étant donné que la structure de fiche est totalement libre, et que ses champs, de longueur variable, n'étant pas limités dans les mêmes proportions que ceux de Foxbase, autorisent un traitement assez souple de données textuelles comme la définition ou le contexte.

Tous les champs peuvent être indexés, ainsi que des articles au sein de ces champs. C'est l'utilisateur qui en décide quand il établit la structure de sa base.

Des procédures de conversion, élaborées par des utilisateurs, permettent désormais d'importer et d'exporter des données de et vers Isis et dBase.

Ce logiciel fait figure de référence en matière de documentation en France. Il est en effet très largement utilisé et il s'est même constitué, chose appréciable, un club des utilisateurs de Texto.

On lui reprochera de se situer dans le haut de la gamme des prix, sans pour autant permettre une consultation à partir d'un traitement de texte (ce qui n'est pas sa vocation), ni une édition totalement satisfaisante sur le plan typographique, si elle l'est presque sur le plan de la mise en page.

Utilisateurs : INRA, Université nouvelle de Lisbonne.

Edibase

Il s'agit encore d'un logiciel documentaire, son originalité étant la possibilité qu'il offre de travailler en texte libre, de procéder à des interrogations en langue naturelle et d'obtenir un tableau de pertinence des résultats. On peut bien sûr utiliser des fiches, si on le désire, mais rien n'empêche de saisir des documents dans leur intégralité. Il comporte, contrairement à Texto, un module guidé d'importation et d'exportation de données. Il faut mentionner également un manuel très satisfaisant pour le néophyte.

Il semblerait également qu'Edibase intègre, dans sa dernière version, un module de repérage de termes, intégralement géré par l'utilisateur.

L'honnêteté nous oblige à avouer que nous n'avons pas encore exploré la totalité de ses possibilités, mais il semble très puissant et a en tout cas paru suffisamment

intéressant à l'Association des conseils en gestion linguistique, basée à Montréal, pour qu'elle lui consacre l'élaboration d'une « trousse » terminologique.

Utilisateurs : ACGL, Radio Canada, Téléglobe, Air Canada.

Cds-Isis

Encore un logiciel documentaire, développé par l'Unesco à Paris. Lui aussi très puissant, on ne peut que lui reprocher toutefois un côté inachevé, qui se caractérise par un manuel très insuffisant, qui rend sa découverte très malaisée, et par sa modularité, qui impose de définir successivement des formats de champs, de saisie, d'édition, d'impression, etc.

Il permet de choisir la langue d'environnement, des menus : français, anglais ou espagnol

L'indexation est possible selon quantité de critères : champs, sous-champs, tous les mots d'un champ, mots signalés par des délimiteurs. La recherche peut s'y faire au choix dans un champ, dans plusieurs champs ou sur toute la fiche, avec historique de recherche, recherche combinée, etc.

Il faut signaler également qu'un module existe, encore à l'état expérimental, de consultation de la base terminologique à partir d'un traitement de texte, et de découpage et de transfert des données de l'un dans l'autre. D'autre part, l'association INFOTERM souhaite poursuivre le développement d'Isis, et pallier en particulier sa modularité pour en faire un véritable outil de traitement terminologique.

Signalons qu'une association des utilisateurs d'Isis vient de se constituer en France et rappelons l'atout majeur d'Isis : il est diffusé gratuitement auprès des organimes publics.

Utilisateurs : Unesco, OCDE, Infoterm, Université nouvelle de Lisbonne.

Les critères à prendre en compte pour le choix d'un logiciel de terminologie

Il est bien évident que le premier des critères à examiner est celui des besoins que l'on éprouve réellement, conjointement avec les moyens dont on dispose.

Veut-on avant tout interroger sa base de données à partir de son traitement de texte ou se satisfait-on de simplement consulter sa base, même au prix d'un passage par Desqview ou pire, par le Dos? Souhaite-t-on procéder à des recherches multicritères ou suffit-il d'une recherche par chaîne de caractères? Est-on assez discipliné pour respecter une structure de fiche imposée ou réclame-t-on toute liberté de structure? Souhaite-t-on publier sous forme dictionnairique ses données terminologiques? A-t-on besoin d'importer ou d'exporter fréquemment des fichiers en provenance des logiciels les plus divers? Désire-t-on pouvoir faire un maximum de choses avec le même logiciel ou envisage-t-on d'en utiliser plusieurs en parallèle?

Ce n'est qu'après avoir procédé à cet examen de conscience, après avoir fait le point sur sa propre pratique, que l'on peut reprendre l'examen des logiciels en tenant compte des éléments suivants, dont la liste n'est sans doute pas exhaustive mais tente de tracer les grandes lignes du logiciel terminologique idéal et multiusage, encore à venir, et dont il faut arriver à trouver le compromis le plus adapté à sa propre vision des choses. On peut ainsi souhaiter disposer des possibilités suivantes :

- structure de fiche libre;
- plusieurs structures de fiche définissables en parallèle et utilisables à volonté;
- procédure d'exportation, guidée, simple, vers le plus grand nombre de logiciels existants;
- procédure d'importation, avec les mêmes critères;
- longueur de champ variable;
- indexation selon plusieurs critères;
- accès aux fiches à partir de l'index;
- recherche multicritère;
- recherche par chaîne de caractères;
- interrogation en langue naturelle;
- création de liens de renvoi de fiche à fiche;
- consultation à partir d'un traitement de texte, avec fenêtrage;
- importation de données terminologiques dans le traitement de texte;
- importation de termes à partir du traitement de texte;
- fonctionnalités avancées de traitement de texte en vue de l'édition (mise en page, polices);
- facilité de transfert vers un traitement de texte;
- aide en ligne;
- contrôle automatique ou manuel des données à partir de fichiers accessibles dans une fenêtre;
- couplage avec un correcteur orthographique;
- saisie possible par scanneur;
- inclusion de données graphiques;
- consultation de dictionnaires du commerce sur support informatique;
- consultation de CD-ROM et importation de leurs données;
- consultation de banques de données externes et importation de leurs données;
- couplage avec un logiciel de repérage automatique de termes;
- couplage avec un lemmatiseur;
- couplage avec un concordancier;
- rapidité de traitement.

Il est bien évident qu'aucun logiciel du marché ne propose toutes ces options, dont bon nombre n'ont pas été évoquées jusqu'ici, quoique leur utilité soit indéniable. Bien qu'il soit fort improbable qu'un tel logiciel existe jamais, on peut toujours souhaiter que les logiciels actuels finissent par intégrer le plus grand nombre des fonctions énumérées ci-dessus et que dans l'intervalle, chacun d'entre nous puisse quand même parvenir à trouver le logiciel le plus adapté possible à ses besoins.

Bibliographie

BÉDARD Claude (1990) « Quoi de neuf en traductique? », *Circuit*, n° 30, septembre 1990.

MAURICE Nathalie (1989) *Évaluation des logiciels MC4, Aquila et Foxbase + dans une perspective terminologique*, mémoire présenté pour le DESS « Information et documentation », Institut d'études politiques de Paris.

MAURICE Nathalie, BLANCHON Élisabeth, OTMAN Gabriel, BOISSY Jacques (1991) « Comparaison de trois logiciels utilisables en terminologie : Foxbase +, MC4, Texto », *Meta*, vol. 36, n° 1.

MAYER Felix (1990) « Terminologieverwaltungssysteme », *Lebende Sprachen*, n° 3, p. 106-114.

DE SCHAETZEN Caroline (1990) « L'ordinateur peut-il fabriquer des dictionnaires », *La Banque des Mots*, n° 40, p. 21-34.

DE SCHAETZEN Caroline (1990) « Outils de bureautique et de télématique pour la traduction », *Lebende Sprachen*, n° 3, p. 123-128.

Répertoire des produits et services de traitement automatique de la langue française, (1989) Observatoire des industries de la langue, Paris, Éditions Daicadif.

Terminogramme, n° 55, hiver 1990, numéro spécial « Industries de la langue ».

23

Stratégies de professionnalisation de la formation des traducteurs

Daniel GOUADEC

Formation des traducteurs, UFR de Langues appliquées, Université de Rennes 2, Rennes, France

Tout programme de formation de traducteurs entend former de véritables professionnels de la traduction. Les moyens utilisés pour y parvenir varient selon la perception que l'on a des fonctions des traducteurs, selon les contextes de formation, et selon les moyens disponibles. Nous présenterons ici, après avoir brièvement analysé les défis auxquels elles répondent et le contexte de leur mise en œuvre, les stratégies retenues dans le cadre du programme de formation de traducteurs à l'Université de Rennes 2.

Les défis

Les contraintes du marché de l'emploi sont, ou sont devenues telles que le traducteur doit nécessairement, au terme de sa formation initiale, être 1) opérationnel, 2) polyvalent (au sens de la multicompétence confirmée) et 3) maître du clavier et de la « bidouille » informatique.

Traducteur opérationnel

Quatre facteurs principaux imposent aux traducteurs d'être capables d'autonomie opérationnelle immédiate. Ce sont le renforcement des exigences de qualité, le coût prohibitif des formations de terrain, la mise hors-circuit des débutants sur appels d'offres, et, bien évidemment, le développement des technologies liées à l'exercice de la profession.

Les exigences de qualité et de productivité ont, conjointement, fait remonter le niveau intrinsèque des premiers recrutements de traducteurs d'entreprises ou d'agences en même temps qu'elles augmentaient l'éventail des « qualités » requises. À compétence égale en traduction, la priorité va toujours à quiconque sait déjà « gérer » ses activités ou même celles des autres : le coût des encadrements de débutants est devenu prohibitif et les cadences de travail interdisent pratiquement toute prise en charge réelle d'un nouveau recruté. Il faut donc que les formations initiales se substituent au moins partiellement aux formations de terrain.

Le traducteur libéral débutant n'est guère mieux loti : s'il ne maîtrise pas intégralement la chaîne de gestion de projets, il n'a guère de chance de décrocher des marchés juteux et risque, au moins temporairement, d'être voué à la sous-traitance de sous-traitance.

Et, bien entendu, la composante technicienne de l'activité de traduction vient compliquer les choses. Les critères de recrutement tendent à glisser progressivement et l'on recrute de plus en plus un traducteur-rédacteur capable d'utiliser tel logiciel ou tel matériel ou de concevoir tel ou tel type d'outil.

Traducteur polyvalent

La polyvalence du traducteur « opérationnel » doit être une réelle multicompétence. Pour reprendre une liste non limitative que tout le monde connaît bien, le traducteur doit maîtriser toutes les formes de traduction, toutes les formes de constitution, gestion, diffusion ou exploitation des terminologies, l'ensemble des techniques de rédaction et de documentation, l'essentiel des principes de révision, sans oublier, bien entendu, tout matériel ou logiciel « imposé » par son donneur d'ordre. Dans l'idéal, le marché recherche toujours un spécialiste du génie linguistique doublé d'un « opérateur » de haut vol.

Traducteur opérateur

Les contraintes de productivité et les jeux de la concurrence ont très largement contribué à mécaniser le lieu de travail du traducteur. Ses outils de travail incluent un logiciel de gestion de fichiers ou de bases de données terminologiques, un ou plusieurs traitements de texte, un ou plusieurs logiciels documentaires et, pour faire bonne mesure, un ou plusieurs systèmes de PAO (publication assistée par ordinateur), sans oublier les mille et un logiciels d'aide au traducteur. Bref, le traducteur doit être prêt à prendre les commandes d'un poste de travail « intégré » et l'ingénierie du traitement des données linguistiques doit devenir une composante majeure des formations.

Le contexte

Conditions générales

Les stratégies de professionnalisation décrites ici ont été élaborées et mises en œuvre dans un programme universitaire de formation de traducteurs recrutant au niveau de chacune des années du premier cycle et au niveau du second cycle, ainsi que l'impose la loi, sans la moindre sélection à l'entrée et sans le moindre test de compétence. Les effectifs sont de 40 étudiants à Bac + 1 (initiation), 30 à Bac + 2 (début de spécialisation des cursus de traduction), 25 à Bac + 3 et Bac + 4 (cursus spécialisé) et 15 à Bac + 5 (DESS d'origines diverses).

Personnel

La « professionnalisation » de la formation est née de la conjonction d'une volonté de relever les défis de l'insertion professionnelle et d'une situation de carence.

Il nous est apparu peu rationnel de créer une formation qui conduirait simplement à mettre sur le marché la copie conforme des étudiants des écoles ou centres existants. Il nous fallait définir des spécificités. En même temps, le très faible encadrement nous interdisait de mettre en place une structure pédagogique complète définie en fonction de cours standard. La seule solution consistait à créer une structure de formation par objectifs professionnalisés de telle sorte que celle-ci soit assurée sur trois postes d'enseignants-chercheurs dont un poste d'enseignant d'informatique.

Moyens

Les moyens ne manquent pas. Les cours se déroulent dans un amphithéâtre avec poste de travail d'enseignant comportant un micro-ordinateur (compatible IBM — disquettes 3,5″ — en réseau sur serveur de programmes) avec projection en continu sur grand écran (système « Barco^{MD} »). Les travaux dirigés se déroulent dans trois salles comptant chacune huit micro-ordinateurs constituant autant de postes de travail d'étudiants, plus une salle comportant douze terminaux. Les micro-ordinateurs sont reliés en réseau (10Net et Ethernet) et peuvent émuler les terminaux. Une salle comptant trois micro-ordinateurs complémentaires, deux terminaux et une station VS3100 reliés au VAX8530 est réservée aux travaux de recherche. Tous les appareils ont accès aux bases de données terminologiques locales et nous disposons de RDB, Word4 et 5, Pagemaker, Paradox, MC4, Aquila, Termex, TermTracer, et des quelque dix systèmes de gestion terminologique mis au point à l'université. Les cent micro-ordinateurs de l'Université (majoritairement des compatibles IBM-PC/PS, mais aussi Macintosh) sont à la disposition des étudiants de traduction comme des étudiants d'autres filières. En travaux dirigés, chaque étudiant dispose de son poste de travail. En régime de croisière, chaque étudiant en traduction travaille au moins 20 heures par semaine sur son poste de travail.

En fait, les conditions matérielles et psychologiques de l'innovation étaient réunies dès le départ et l'option de professionnalisation systématique a constitué un atout et non une contrainte. Les trois paramètres déterminants de la professionnalisation ont été la volonté de « coller aux marchés » malgré l'interdiction de toute sélection, la pénurie des encadrements et la présence du parc informatique et des informaticiens dont le renforcement ces dernières années doit beaucoup au développement du programme de formation des traducteurs et du centre de recherche qui s'y rattache.

Les stratégies

Les stratégies de professionnalisation de la formation des traducteurs abordées ici sont au nombre de dix. Elles portent sur le développement de l'information, la définition des champs de formation et des cursus, l'intégration verticale et horizontale des cursus, la définition et la mise en œuvre systématique d'un cahier des charges pour toute tâche à réaliser, la définition d'une progressivité raisonnée des apprentissages, la recherche des indices de l'évolution des compétences (et, notamment, des régressions positives), l'analyse des rendements des diverses stratégies envisagées, la création d'un poste de travail « professionnalisé », la mise en situation de responsabilité, et la prise en compte, par anticipation, de l'évolution des marchés.

Se renseigner et renseigner sur les spécifications du produit

Les formations de traducteurs étant toutes, par vocation, professionnelles, elles doivent atteindre des objectifs conformes aux exigences de la profession (les règles de l'art) et du marché (les exigences justifiées des donneurs d'ordre).

Afin que ces objectifs nous soient connus et qu'ils soient diffusés auprès des étudiants, nous pratiquons une politique d'ouverture systématique à tout traducteur, terminologue, rédacteur, etc. qui accepte de présenter son expérience. Nous considérons surtout que les représentants des entreprises et donneurs d'ordre doivent venir avaliser la formation à l'intérieur même de l'Université, notamment par l'intermédiaire des partenariats de formation. Nous multiplions les présentations, conférences, visites, actions de partenariat et stages divers, qui permettent la mise à jour permanente des spécifications de notre « produit fini ». Ces spécifications étant connues, nous concevons chaque objectif de formation en fonction de compétences confirmées par des travaux « grandeur nature » effectués dans des conditions « vraies » de qualité, de délais, d'outils. La formation se mesure ainsi en lignes de c.v. et notre avantage principal réside dans le fait que la maquette qui définit la formation impose uniquement le respect d'un nombre d'heures de cours. Nous pouvons donc créer, réaménager ou supprimer des enseignements d'une année universitaire à l'autre à la seule condition de ne pas accroître indûment la charge horaire. En pratique, nous introduisons librement de nouvelles activités dans une structure de formation totalement ouverte.

Définir les champs et les cursus

Les spécifications du « produit » étant définies, nous organisons le cursus selon un tronc commun fixant les règles de l'art et un ensemble de composantes aléatoires renvoyant à des créneaux, des pratiques, ou des « avances » spécifiques. Ce sont ces composantes aléatoires qu'il a fallu définir avec soin.

Dans la définition des champs (ou domaines) d'application des compétences, deux grandes questions se posent. La première question concerne la « plus-value » apportée par l'activité du traducteur. Elle conduit à faire un choix entre sur-spécialisation technique du traducteur et polyvalence. La seconde concerne la part de « plus-value » que la formation apporte au traducteur. Elle conduit à s'interroger sur les domaines et spécialisations les plus porteurs d'emploi ou de rémunération.

Les réponses que nous avons données à ces deux questions sont fort simples : nous formons le traducteur de telle sorte qu'il puisse s'intégrer à une équipe à laquelle il apportera une compétence linguistique-langagière hyper-spécialisée appuyée sur des compétences techniques générales. Il reçoit donc, en complément à sa formation spécifique de technicien de la manipulation des langages, une formation technique générale et une formation plus spécialisée en informatique auxquelles il doit ajouter, par autoformation, une sur-spécialisation technique individualisée.

La formation technique générale s'organise au fil d'un ensemble de travaux terminographiques portant obligatoirement sur des secteurs croisés ou composites et conduits sous l'autorité d'un technicien confirmé. La formation en informatique prend la forme de cours et travaux pratiques. L'autoformation s'organise autour des activités documentaires.

Les champs d'application et les modalités générales de la formation « technique » étant ainsi définis, les domaines de compétence langagière d'un traducteur en fin de formation (Bac + 4) incluent obligatoirement :

- la traduction générale,
- les conduites de la traduction spécialisée,
- la rédaction générale et la rédaction de notices,
- la révision de traductions (dont la post-édition),
- la révision de rédactions,
- la conduite des activités de documentation,
- la constitution de données terminologiques,
- la gestion des données terminologiques,
- la totalité des formes d'exploitation des terminologies,
- la gestion des données phraséologiques,
- la gestion et l'amélioration d'un poste de travail intégré.

La question cruciale devient alors la suivante : Comment « intégrer » les divers éléments de formation?

Intégrer les éléments de formation

Trois formules d'intégration sont nécessaires. Il faut d'abord imbriquer les unes dans les autres les activités contribuant à l'acte de traduction qui définit l'axe majeur de la formation (terminologie, documentation, rédaction, etc.). Il faut ensuite intégrer des divers « moments » de formation. Il faut enfin, dans un contexte de faiblesse quantitative des encadrements, intégrer la pyramide des années.

Intégrer les activités constitutives de l'acte de traduction a été facile dès l'instant où nous avons décidé que chaque objectif de formation correspondrait à un produit-traduction donné. Ce produit étant défini (par un cahier des charges), les apports de connaissances, de ressources, d'outils et de procédures relevant de chacune des activités concernées sont effectués au moment voulu et dans des conditions de pratique spécifique. L'intégration exploite le schéma de progressivité des apprentissages qui sera considéré ultérieurement. En fait, le programme se construit tout seul, à condition que la structure pédagogique soit suffisamment souple pour permettre les « cours à la demande ».

L'intégration des divers moments de formation par activité s'effectue sans grande difficulté. Elle repose sur un enchaînement systématique. En un premier temps, un cours unique expose à l'ensemble des étudiants du programme les enjeux, méthodologies et procédures relatifs à l'activité concernée. Ce cours unique exposant les règles, risques et principes conduit à la définition du cahier des charges type. Il est suivi d'une démonstration sur poste de travail d'enseignant, par conduite effective de l'ensemble ou d'une partie de l'activité concernée, avec affichage sur écran géant. Le fichier comportant l'ensemble des données et états de réalisation de la tâche est enregistré sur l'un des disques durs réservés et mis à la disposition de tous les étudiants (polycopie de disquettes). Une ou plusieurs séances de travaux pratiques, par niveau, permettent toutes les mises au point nécessaires en situation de réalisation de la même activité. L'enseignant intervient pour renseigner et réviser. Tout élément significatif est noté sur la disquette de l'étudiant. Lorsqu'une règle émerge, elle fait l'objet d'une directive. Les règles et directives générées en travaux pratiques sont périodiquement fusionnées et le fichier commun normalement copié et étudié par l'ensemble des étudiants de la filière. Parallèlement, un fichier commun des erreurs types se constitue par groupe d'année puis par fusion inter-années et devient accessible à tous les étudiants de la formation.

Lorsque les travaux pratiques confirment que les techniques et procédures sont maîtrisées, les premiers travaux internes-externes débutent. Il s'agit de travaux confiés par des donneurs d'ordre extérieurs à l'Université et réalisés par les étudiants sous le contrôle des enseignants-réviseurs. Les procédures pédagogiques restent les mêmes, à savoir : définition du cahier des charges spécifique, constitution d'un fichier commun de règles de procédures et d'un fichier commun des erreurs types. Le serveur permet en outre l'exploitation d'un système dit de « conférence », grâce auquel chaque traducteur ou rédacteur ou terminologue peut à tout moment poser une question à tous ses partenaires, à l'informaticien responsable du système, au chef de projet, et au réviseur-enseignant. Les réponses sont communiquées à tous et contribuent à l'affi-

nement du cahier des charges. Une fois les travaux internes-externes maîtrisés, les étudiants passent aux travaux externes et aux stages.

Reste l'intégration par années. Une fois encore, les conditions de pénurie ont intégré la pyramide des années en ce sens que les réalisations diverses sont effectuées par des groupes constitués verticalement. Ces groupes incluent donc des étudiants de seconde année (dans un rôle d'observateurs), de troisième année, et de quatrième année. Dans ces groupes, chacun apporte sa contribution dans les limites de ses compétences et en fonction de ses affinités. Le groupe est le véritable creuset de formation : l'inter-révision est la règle et quiconque maîtrise un savoir ou un savoir-faire le transmet à l'ensemble de ses « collaborateurs ».

L'ensemble de la formation repose sur trois règles : une règle de solidarité à l'intérieur du groupe, une règle de généralisation des partenariats d'entreprises, et une règle absolue de transfert de compétences, notamment en retour de stage.

Définition d'un cahier des charges pour chaque activité

Pour chaque réalisation effective (réalisation d'un dictionnaire relatif à la traduction commandée, traduction d'un document, révision d'un lexique, définition d'une structure de fichier terminologique, rédaction d'une notice technique, etc.) un cahier des charges strict doit être établi. Ce cahier des charges :

- précise l'ensemble des directives, normes, et conventions;
- indique les responsabilités de chacun;
- stipule le modèle d'organisation du travail retenu;
- régit la conduite des activités;
- fixe les protocoles d'échange d'information;
- normalise la présentation du produit (feuilles de style);
- fixe les délais et l'échéancier par activité ou sous-produit;
- liste les ressources disponibles ou recommandées ou imposées;
- précise les outils et supports éventuellement imposés;
- définit au besoin les modalités de négociation avec le donneur d'ordre;
- pose les critères de qualité révisable (seuil en-deçà duquel le réviseur refuse légitimement d'intervenir);
- pose les critères de qualité livrable (seuil de qualité finale);
- fixe les conditions de rétribution (notation ou paiement des participants à l'activité).

La définition du cahier des charges par consultation du donneur d'ordre puis réflexion collective constitue en fait le premier acte pédagogique. Il pose le formateur dans sa fonction de personne-ressource chargée d'arbitrer, de conseiller, de guider. Il définit, pour le formateur, des fonctions d'« aide en salle » et de chef d'agence virtuelle.

Définir la progressivité

À ceux qui craindraient de voir le « professionnel » tuer le « pédagogique », on s'empressera de faire valoir que la professionnalisation nourrit la pédagogie. En effet, les

objectifs de qualité obligent à définir des paliers ou niveaux de compétence, aussi bien dans une perspective d'enseignement que dans une perspective de pratique.

Les recherches que nous avons conduites montrent que la formation du traducteur peut s'organiser par strates correspondant à des objectifs cohérents, homogènes et réalisables en totalité, et non pas par cumul de « solutions » à des problèmes parcellisés plus ou moins aisément répertoriables. Nous définissons donc la progressivité des apprentissages selon un axe central de types de traduction enchâssés : traduction signalétique, traduction sélective-documentaire, traductions synoptiques, traductions banalisées et traductions absolues. Chacun de ces types correspond à un palier formant la condition nécessaire du passage au type suivant et chacun d'entre eux correspond à un « produit » dont la validité est incontestable dans l'économie du marché de la traduction. En tout état de cause, chacun d'entre eux correspond à un objectif pédagogique parfait et parfaitement cohérent.

Chaque type de traduction repose sur un cahier des charges précis et mobilise un niveau clairement défini de compétence pour chacune des activités contributives (compréhension, terminologie, rédaction, etc.). Le passage d'un type (niveau) à l'autre s'effectue par glissement et non par saut qualitatif extrême, et l'étudiant n'est jamais amené à dépasser son niveau de compétence.

Soit, ainsi, la traduction sélective-documentaire qui doit fournir, sous une forme simplifiée, la seule information « utile » à un donneur d'ordre particulier, sans obligation de rédaction d'un texte, mais en veillant à ce que l'information soit parfaitement transparente et en incluant donc toute explication nécessaire. Ce type de traduction requiert une parfaite maîtrise de la délimitation des cadres d'information, un repérage des trames d'objets, un repérage systématique des segments « utiles », un réaménagement des relations thématiques, une réorganisation de la structure profonde des données, la construction d'un index documentaire sans faille, une recherche documentaire complète, la mise en place de tout élément capable d'annuler les écarts culturels, le traitement complet et en contexte de l'ensemble des termes spécialisés. Pour simplifier à l'extrême, la traduction sélective-documentaire annule les effets de rédaction, trie l'information, met en place des ponts avec l'univers de référence. Elle réduit les objectifs et les contraintes qui sont ceux de la traduction totale et totalement textualisée mais conserve leur caractère absolu à ceux des sous-objectifs qu'elle pérennise. Elle garantit surtout que les bases sur lesquelles reposera le type suivant sont justes.

Le principe de progressivité trouve une autre application générale : chacun doit s'arrêter dès l'instant où sa compétence elle-même s'arrête. Ceci constitue la règle de fonctionnement des groupes de travail. Chacun apporte sa pierre à l'édifice commun et un étudiant de seconde année peut être appelé à définir l'index documentaire ou à effectuer un inventaire terminologique transmis pour traitement aux étudiants de troisième année avant révision par des étudiants de quatrième année.

Repérer les indices de progression

Les passages d'un type-niveau de traduction à l'autre et, donc, d'un état de compétence composite à l'autre, constituent la clé du système. La professionnalisation des travaux,

leur volume considérable (il est courant qu'un groupe de 5 étudiants ait à traduire un texte de 80 pages en 10 jours), et le principe de révision systématique par le formateur accélèrent le repérage des indices de l'évolution des compétences.

Le cahier des charges énumère les différentes sous-tâches constitutives de la tâche de traduction (ex. saisie de l'original, repérage des récurrences, extraction des termes spécialisés, repérage des stéréotypies, index documentaire, liste des hybrides, recherche des « équivalents », notes documentaires, échantillons, traduction, pointage, contrôle d'homogénéité, contrôle de lisibilité, vérifications techniques, essais, validation) avec, pour chacune, un ensemble de directives. L'affectation des tâches s'effectue en fonction des niveaux atteints et des désirs de spécialisation des uns et des autres. Le principe de responsabilité collective fait que tout sujet recevant une « contribution » d'un autre sujet en vérifie nécessairement la qualité. Ce sont donc les membres d'une même équipe de production qui définissent les compatibilités de niveaux et les indices de progression en référence aux données du cahier des charges : une « pièce » contribuant à l'ensemble n'est acceptée que si elle répond aux critères de la qualité « livrable » au niveau suivant. La vérification est de type ascendant par contrôle de qualité des composants et de type descendant par retour après chaque niveau subséquent et, bien entendu, après le niveau final de révision par le formateur ou le donneur d'ordre.

La révision systématique des travaux permet de faire apparaître des lignes de force, de connaître parfaitement chacun des sujets, et de suivre les progressions des uns et des autres. En pratique, nous nous attachons à deux types de phénomène qui sont, d'une part, les erreurs types et, d'autre part, les indices significatifs. Les erreurs types sont recensées et signalées à l'ensemble du groupe concerné, puis de tous les groupes, toute nouvelle occurrence postérieurement à leur signalisation étant considérée comme faute fatale. Les indices significatifs sont les indices de progression (état neutre) et les indices de régression « positive ». Ces derniers témoignent d'un changement d'état ou de comportement apparaissant en précurseurs de sauts quantitatifs et dus, selon toute vraisemblance, à une déstabilisation de comportements acquis. Ils signalent des moments critiques dans la mesure où ils correspondent à des phases dans lesquelles tout (y compris les comportements aberrants) peut se mettre en place mais aussi dans lesquelles le formateur doit intervenir pour guider, expliquer, rassurer. On peut citer, à titre d'exemple, les régressions « positives » inévitables que constituent le surclassement des terminologies dans la traduction signalétique, les ruptures pronominales dans la traduction sélective, la déstabilisation syntaxique dans la traduction banalisée, la divergence systématique dans les premières phases de la traduction absolue, etc.

Les indices de progression/régression temporaire déterminent la chronologie des interventions personnalisées. Celles-ci portent un peu sur les stratégies génériques de la traduction et beaucoup sur les stratégies spécifiques liées à des particularités de domaines ou de types d'activités. La relation entre le traducteur en formation et le formateur passe par l'objet à traiter et non plus par le cours sur le traitement de l'objet.

Analyser les rendements des stratégies / comportements / procédures

Suivre la progression générale d'un groupe et les progressions particulières de chacun des membres du groupe ne suffit pas. Il faut s'interroger sur le rendement des diverses

procédures et stratégies dans des conditions tendues : délais particulièrement raccourcis, terminologie imposée par le donneur d'ordre, terminologie imposée par des traductions antérieures, traduction, dans un ordre aléatoire, de fragments ou sections d'un même document, illustrations nombreuses, contraintes de limites de nombre de caractères, etc. Il s'agit d'analyser, entre autres, les conditions et effets de la division du travail, de la traduction par couches, de la traduction dictée, de la post-édition. Former des traducteurs, c'est aussi les préparer aux situations particulières et particulièrement tendues qu'ils risquent de rencontrer dans leur future vie professionnelle. L'expérience prouve que, dans ce type de situation (habituellement) exceptionnelle, tous les acquis basculent et tous les comportements aberrants réapparaissent. Mais c'est aussi dans ce type de situations que l'on peut sceller définitivement les bonnes habitudes. Les « ratages » splendides induits par la mise en place de conditions de tension ont une vertu capitale : ils démontrent la faillite de comportements ou stratégies d'apparence logique et il n'est alors nul besoin de discours pour mettre en place des stratégies efficaces.

Créer le poste de travail

Il n'est pas de professionnalisation réelle sans le recours aux outils professionnels. Nous utilisons donc un poste de travail. Mieux, chaque étudiant crée son poste de travail.

Chacun commence, au fil de ses activités diverses, par définir les fonctionnalités souhaitées au-delà du traitement de texte, de l'accès aux bases existantes, et de toutes les fonctions standard. En règle générale, les demandes portent sur la mise en place automatique d'une gestion d'écrans par type de tâche, l'importation de graphisme et de données, la consultation des dictionnaires en ligne, l'extraction automatique des termes, la constitution d'un inventaire des blocs récurrents, la recherche automatique des contextes en fonction de listes d'indices spécifiques, le tri des inventaires sur contenus d'anti-dictionnaires, la définition des indices contextuels positifs, la génération automatique de demandes de documents, la génération automatique de données terminographiques, les substitutions terminologiques, les substitutions phraséologiques, les importations de structures de formulaires, et toutes les formes de guide.

Une fois définies les fonctionnalités souhaitées, chacun élabore l'algorithme de mise en place. Cette élaboration relève du travail pratique permanent en informatique.

Les différentes composantes du poste de travail sont ensuite élaborées (développées) individuellement ou par groupes, avec, bien entendu, bourse d'échanges finale. La progression de mise en place est toujours la même : simulation en traitement de texte, émulation en traitement de texte, développement de système dédié.

À l'Université de Rennes 2, nous avons deux *a priori* absolus. Le premier dit que le traducteur traite principalement du texte, que le logiciel le plus radicalement interfacé avec le traitement de texte est ce traitement de texte lui-même, et que l'on peut tout faire au stade artisanal avec le traitement de texte. Le second dit que la puissance de calcul, c'est l'affaire des gros calculateurs et que le passage au stade non artisanal

suppose la mise en œuvre d'un gros système. Entre les deux il n'y a que du vent commercial.

Or donc, chaque étudiant commence par simuler, pas à pas, manipulation par manipulation, la fonction qu'il souhaite créer (ex. chercher automatiquement dans plusieurs fichiers existants tous les contextes = phrases dans lesquels se trouve la chaîne de caractères xxx, regrouper tous les contextes dans un même fichier, chercher dans ce fichier tous les contextes contenant un indice certain de présence d'une définition, et ainsi de suite).

Une fois la simulation maîtrisée et la nature des divers problèmes posés clairement dégagée, l'étudiant écrit la macro-commande qui réalisera automatiquement la fonction en traitement de texte. En pratique, nous disposons actuellement d'une cinquantaine de macros de traitement de texte allant de celle qui prépare automatiquement l'écran pour la traduction, avec fenêtre du texte à traduire, fenêtre de la traduction, fenêtre de consultation des dictionnaires, fenêtre d'aide-mémoire, etc. à celle qui remplace automatiquement dans un texte les légendes des figures en passant par celle qui propose des schémas standard de démonstrations mathématiques et par celle qui met en évidence les principes et limites de la traduction automatique. L'imagination a pris le pouvoir et il n'y a rien de plus pédagogique que de tenter de mécaniser ou d'automatiser un processus.

En bout de course, l'étudiant définit un cahier des charges destiné à servir de guide à la mise en place de la fonction considérée sur le gros système et sur les stations de travail multifenêtrées.

Mettre en situation de responsabilité

Chacune des activités, sous-activités ou composantes d'activités précédemment considérées engage des responsabilités particulières des étudiants.

Quelle que soit son année d'études, chaque étudiant participe à la totalité des activités de projets : traduction, rédaction, révision de traduction, révision de rédaction, révision de travaux terminologiques, terminographie, saisie, documentations, constitution des bases de données phraséologiques, réalisation de macro-commandes. À ce titre, et quelles que soient ses attributions spécifiques par ailleurs, il applique les directives précises du cahier des charges et rend compte au chef de groupe (ce dernier ayant défini les seuils de qualité).

À Bac + 3 ou au-delà, chaque étudiant remplit, pour trois projets au moins au cours de l'année universitaire (un projet terminologique, un projet documentaire et un troisième projet au choix), les fonctions de chef de groupe. Il doit alors faire appliquer les directives du cahier des charges, fixer les seuils de qualité à partir desquels il intervient en réviseur, faire appliquer les directives et critères, et rendre compte au chef de projet.

À Bac + 4 ou au-delà, chaque étudiant remplit au moins une fois les fonctions de chef de projet. À ce titre, il définit le cahier des charges en accord avec le donneur d'ordre, élabore les directives qu'il transmet aux chefs des divers groupes participant

à son projet, fixe les seuils de qualité, fait appliquer les directives et critères. Il rend compte au formateur responsable et au donneur d'ordre.

La structure de responsabilité ci-dessus constitue la colonne vertébrale du système. Elle contribue à mettre en place des principes de gestion de projets aux termes desquels tout est prévu, puis vérifié, dans le moindre détail. La règle, impitoyable, dit qu'un travail est refusé tant qu'il ne répond pas aux critères de qualité imposés. Elle dit aussi que le chef de projet ou de groupe est responsable de toute carence due à une absence, une ambiguïté ou une imprécision des directives. Elle contribue à généraliser l'esprit de collaboration.

Devancer les évolutions du marché

Dans la perspective que nous avons retenue, les lignes de force du marché sont implicitement présentes en permanence. Elles le sont dans la définition de types de traduction se démarquant du « tout traduit ». Elles le sont dans la mise en place de types variés de gestion des traductions. Elles le sont aussi dans la recherche des accroissements de productivité liés à l'amélioration des stratégies individuelles ou collectives mais aussi et surtout au développement des aides au traducteur, à la répartition des tâches, au recours à la traduction dictée. Elles le sont enfin dans la prise en compte de nouveaux types de pratique et, notamment, dans l'avènement — peut-être regrettable — de la post-édition.

Mais il ne suffit sans doute pas de suivre les évolutions rapides des conditions d'exercice de la profession. Il est utile de tenter de les prévoir et de mettre déjà en place les bases de la future formation continue des traducteurs. Nous avons, à cet égard, choisi trois voies qui, à des degrés divers, nous paraissent prometteuses. Ce sont, par ordre d'importance croissante, la constitution de banques de données phraséologiques, la définition d'un poste de travail multifonction et l'informatisation des traducteurs.

La première voie nous amène à réfléchir sur les stéréotypies nationales en traduction et débouche sur la mise en place de canevas naturels de traduction et rédaction. La seconde voie prend appui sur un principe de détournement du logiciel « naturel » des traducteurs (traitement de texte sur micro-ordinateur) pour en faire un outil de travail et un outil de formation capable de tout. La troisième voie conditionne tout le reste mais, pour que les choses aillent dans le bon sens, nous pensons qu'il faut donner aux traducteurs les moyens de prendre le pouvoir informatique.

Conclusion

Aussi étrange que cela puisse paraître, l'accès à un parc de matériel abondant ne conditionne nullement tout ce qui précède. Il faut en fait considérer qu'il existe deux options.

L'option maximale correspond à la formation de traducteurs appelés à exploiter des matériels divers et à travailler dans un univers lourdement informatisé. En pareil cas, le centre de formation doit disposer d'un environnement comparable et y préparer, mais il ne s'agit que d'une plus-value par rapport à la seconde option.

L'option minimale correspond à une situation de carence en matériel et formateurs en informatique. En pareil cas, on fait abstraction de l'outil et de la compétence particulière, mais rien n'invalide la philosophie générale du système.

Il faut considérer, en tout état de cause, que l'ensemble a été conçu pour faire face à la criante pénurie de formateurs. Il permet à un minimum d'enseignants-réviseurs de gérer la formation et les travaux en terminologie, traduction, rédaction, etc. d'un nombre important d'étudiants. L'option d'encadrement de travaux effectifs constitue la seule réponse possible dans la mesure où elle sollicite chacun à son niveau dans une perspective qui reste définie par un objectif de réalisation de tâche complète. Elle fait intervenir une définition d'objectifs par directives, un encadrement permanent des activités, un suivi en continu des évolutions, une correction quasi immédiate, une prise en charge des comportements. Elle permet surtout de déléguer en cascade une bonne part de la formation par transfert-transmission de savoirs et savoir-faire à l'intérieur des groupes. Elle conduit à mettre en œuvre une structure de responsabilités. Elle importe dans l'institution universitaire, en les réévaluant au besoin, les modes de fonctionnement, contraintes, pratiques, outils, et responsables des services de traduction. Elle reste viable quand bien même elle n'est plus assistée que par gomme et crayon.

24

Quelques logiciels utiles aux traducteurs ou aux terminologues. Les hypertextes et l'analyse des contextes

Charles DOUTRELEPONT

Département de français, Université de Carleton, Ottawa, Canada

Il y a fort longtemps, au XIII^e siècle, un cardinal dominicain, Hugues de Saint-Cher, de Paris, eut l'idée de recenser tous les mots qui figuraient dans la Bible[1]. Hughes souhaitait disposer d'une liste complète des mots de la Bible, plus précisément ceux de la Vulgate, version médiévale de l'Écriture sainte. Près de 500 moines, sous la direction de Hughes de Saint-Cher, ont ainsi répertorié et classé par ordre alphabétique tous les mots de la Vulgate du XIII^e siècle. Dès le départ, le projet fut sensiblement modifié : Hughes ne souhaitait pas disposer uniquement d'une liste de mots, mais aussi d'un index. Commentateur de la Bible, ce dominicain voulait pouvoir retrouver les différents contextes dans lesquels tel et tel mot s'observe. Dès le XIII^e siècle, on a ainsi disposé d'index bibliques, c'est-à-dire de listes de mots au sein desquelles chaque mot est accompagné de sa ou ses références au texte. Par index, il faut entendre un inventaire des mots-formes, si l'on veut être précis, puisqu'un mot qui figure plus d'une fois dans la Bible, par exemple, n'apparaît qu'une seule fois dans l'index qui y est rattaché. Pour ce qui est des références, chaque mot est associé à l'ensemble des références qui renvoient aux différents passages de la Bible au sein desquels le mot apparaît. À partir de cette liste de mots accompagnés de leur référence, il est possible

1. Les éléments qui concernent les concordances de la Bible sont empruntés à J. Cochrane et G. Chouinard, *Concordance et Index de la Bible, Tome II, Nouveau Testament*, Distribution évangélique du Québec, 1987, p. III.

de savoir où les mots se manifestent dans le texte. Mais pour étudier ces mots dans leurs différents contextes, il demeure nécessaire de parcourir le texte dans un mouvement allant d'un contexte à l'autre et dans un va-et-vient répété de l'index aux contextes.

Au XVI[e] siècle, on élabore un outil plus sophistiqué, la première concordance. Comme dans un index, tous les mots du texte figurent dans une concordance mais, contrairement à l'index, la concordance rassemble toutes les occurrences des mots d'un texte. C'est donc dire que, si un mot s'observe sept fois dans un texte, il est repris sept fois dans la concordance. Contrairement à l'index aussi, chaque occurrence prend place au sein de son contexte original, celui du texte. La disposition traditionnelle d'une concordance consiste à regrouper toutes les occurrences d'un même mot sous une rubrique, un mot-vedette. Sous cette rubrique, toutes les occurrences du mot apparaissent, en général à raison d'une occurrence par ligne, dans leur contexte respectif. Par rapport à l'index, la concordance rapproche les contextes dans lesquels s'observent les mots : elle dispose ces contextes les uns à la suite des autres, et elle favorise, par conséquent, les comparaisons de tous ordres.

La concordance comporte cependant certaines limites; les contextes y sont le plus souvent limités à quelques lignes et, surtout, la dimension de ces contextes y est constante du début jusqu'à la fin. Pour pallier cette lacune, chaque occurrence est encore pourvue d'une référence au texte, d'un renvoi, qui permet, au besoin, de retourner au texte original et d'obtenir un contexte plus large que celui que fournit la concordance.

La première concordance française, la concordance de la Bible, basée sur les versions Segond et Synodale de l'Écriture, a été publiée en 1566 à Genève, vraisemblablement sous l'influence de Calvin.

Index, concordance, ces outils, désormais produits par ordinateur, demeurent encore aujourd'hui des instruments utiles pour ceux qui souhaitent explorer les mots dans leurs différents contextes. Dans les pages qui suivent, nous présenterons les logiciels qui permettent cette exploration.

Deux types de logiciels sont actuellement disponibles pour l'étude des mots en contexte : ceux qui ont été conçus pour le traitement en lot et ceux destinés à une utilisation en mode conversationnel. Le premier type de logiciels, employé depuis une quarantaine d'années, permet de produire des index et des concordances sur papier. Le deuxième type, plus récent, autorise une exploration des contextes à partir d'un dispositif visuel, d'un écran cathodique; cet instrument de travail d'un type nouveau est beaucoup plus souple et offre de nouvelles possibilités aux chercheurs.

Les logiciels de traitement en lot qui permettent de générer concordances et index sont bien connus des chercheurs. À partir d'un texte enregistré sur un support informatique, disque ou disquette, rubans ou bandes magnétiques, l'ordinateur découpe le texte en mots. Un mot y est défini comme toute suite de caractères comprise entre deux espaces blancs, deux signes de ponctuation ou un espace blanc et un signe de ponctuation. Dans le cas des index, les différentes occurrences d'un mot sont accompagnées de leurs références au texte ou, en langage informatique, de leurs « adresses »,

alors que dans le cas d'une concordance, les différentes occurrences d'un mot sont imprimées dans leur contexte.

Sur le plan des réalisations, les index informatisés ont précédé les concordances. En 1949, le père R. Busa[2], à Gallarate, en Italie, a entrepris la saisie sur cartes perforées de la *Somme théologique,* de saint Thomas d'Aquin, l'*Index Thomisticus*. Dès les débuts de l'entreprise, l'*Index Thomisticus* n'était pas un index au sens strict, complété qu'il était par la concordance des œuvres. À partir de 1960, de nombreuses concordances ont été publiées et des progiciels ont été mis au point pour en faciliter l'élaboration : JEUDEMO[3], à l'Université de Montréal, OCP[4], l'Oxford Concordance Program, à Oxford. Conçus au départ pour de gros ordinateurs universels, certains de ces progiciels sont aujourd'hui disponibles sur micro-ordinateur; c'est le cas de l'Oxford Concordance Program. Des logiciels de traitement de texte, comme Nota Bene, permettent aussi d'envisager la production d'index ou de concordances.

Dans tous les cas, les concordances produites à partir de tels logiciels demeurent des concordances conçues pour être imprimées. Dans le cas des logiciels conçus pour effectuer un traitement en lot, le papier demeure donc un support essentiel. C'est sur papier que les résultats sont disponibles au terme du traitement. Cette considération n'est pas sans conséquence.

Facteur difficilement quantifiable mais essentiel, l'information qu'il faut considérer est multipliée par 10 ou par 20. Par exemple, dans le cas d'un document de 200 pages, une concordance imprimée de ce document fera, à raison de 10 mots de contexte par mot concordé, près de 2000 pages. Si l'on souhaite avoir 20 mots de contexte, on atteint vite une masse considérable de 4000 pages, et cela toujours pour un document original de 200 pages. Les chercheurs qui ont eu recours à des instruments de travail de ce type connaissent les inconvénients des concordances imprimées sur les imprimantes des centres de calcul : manipulation délicate d'une masse de papier de mauvaise qualité et reliure médiocre, si tant est que l'on puisse relier l'ensemble. De plus, le document ainsi produit s'use très rapidement. Quand ces concordances sont publiées et diffusées dans le commerce, leur prix est souvent élevé, puisqu'il est directement relié à la quantité de papier imprimé.

Ces concordances posent encore un autre problème, celui de la dimension des contextes, qui est fixe et arbitraire. Le nombre de mots réservés au contexte, à gauche et à droite du mot-vedette, est parfois trop peu élevé, parfois trop élevé. Pour certains mots, il est souhaitable de disposer de deux ou trois mots de contexte, tandis que pour d'autres, il faut avoir deux paragraphes de contexte. Il est évidemment possible de changer les paramètres d'impression des concordances, d'augmenter ou de diminuer, par exemple, le nombre de mots souhaités. Mais ces changements supposent non

2. Busa S.J. R. (1976) « Computer Processing of Over Ten Millions Words : Retrospective Criticism », A. Jones and R. F. Churchouse (eds.), *The Computer in Literary and Linguistics Studies*, Cardiff, University of Wales Press, p. 114-17.
3. Bratley P., S. Lusignan, F. Ouellette (1974) « JEUDEMO : A Text-Handling System », *Computers in the Humanities*, p. 234-49.
4. Hockey S. (1985) « Literature and the Computer at Oxford University », p. 53-79, B. Derval et M. Lenoble (éd.), *La critique littéraire et l'ordinateur*, Montréal.

seulement que l'on modifie les paramètres de la concordance, mais aussi que l'on reprenne l'exécution complète du programme. Les logiciels de ce type sont conçus pour un traitement en lot, en différé, ce qui a pour conséquence que l'utilisateur du programme ne peut intervenir à partir du moment où le programme est lancé. Toute modification des paramètres entraîne une nouvelle exécution du programme.

Les logiciels conçus pour un traitement en mode conversationnel, les « langages de programmation orientés objet » que l'on appelle « hypertextes », permettent au contraire à l'utilisateur d'intervenir en cours de traitement. À partir de tels logiciels, il est possible, par exemple, de modifier en tout temps les paramètres de présentation d'une concordance, celle-ci étant explorée d'abord à partir de l'écran. C'est là un des avantages de ce deuxième type de logiciels. Nous allons maintenant les aborder en détail, en examinant, à titre d'exemple, le fonctionnement, les caractéristiques et les possibilités de TACT. TACT[5] est un système de repérage de l'information qui a été conçu à Toronto, par les membres du Centre for Computing in the Humanities. Au Canada, TACT est distribué gratuitement aux institutions universitaires membres du Consortium pour les ordinateurs en sciences humaines.

Techniquement, TACT est un hypertexte, c'est-à-dire un programme qui gère des blocs d'information structurés, ou si l'on veut des ensembles d'informations hiérarchisés, délimités. Ces considérations sont fondamentales, car elles peuvent rendre compte de la différence essentielle entre ce type de logiciels et les logiciels conçus pour le traitement en lot.

Les informations gérées par les programmes de traitement en lot sont des informations linéaires, puisque, dans ce système, mot concordé, mot indexé, références et contextes sont disposés sur une seule ligne. À partir de ces informations linéaires, divers tris peuvent être opérés : tri par ordre alphabétique, tri par ordre alphabétique inversé, etc. Contextes et références sont paramétrisables : on peut augmenter ou réduire la taille des contextes, on peut trier les mots selon des critères de fréquence. L'essentiel, c'est la structure linéaire de l'information.

Contrairement aux programmes de traitement en lot, les programmes du type TACT, les hypertextes, gèrent des blocs d'information hiérarchisés et délimités, des « nœuds ». Ces blocs d'information peuvent être constitués d'informations de nature très variée. Il peut s'agir du texte intégral sur lequel porte l'analyse du chercheur; il peut s'agir d'une liste de mots que l'on souhaite étudier dans leurs contextes, il peut s'agir d'options de présentation d'un index ou d'une concordance. Les blocs d'information sont en outre délimités par des symboles qui en indiquent les bornes. Ce peut être des signes qui indiquent le début ou la fin d'un titre, des caractères qui encadrent les données d'un champ d'information d'une fiche terminologique. Tous ces « blocs d'information » ont en commun une caractéristique essentielle : ils sont accessibles en tout temps. Concrètement, ces « blocs d'information » apparaissent à l'écran sous la forme de fenêtres, c'est-à-dire d'encadrés à l'intérieur desquels figure ladite information.

5. Pour en savoir plus sur TACT on peut consulter le guide de l'utilisateur, *TACT, User's Guide, Version 1.1*, Toronto, University of Toronto (Centre for Computing in the Humanities), 1989, 135 p.

Dans le cas de TACT plus précisément, on peut distinguer cinq niveaux hiérarchiques, allant de la base de données textuelles proprement dite aux différents modes de présentation possible, index et concordance, par exemple.

Au niveau inférieur de la structure se situe le texte, ou plus exactement la base de données textuelles. Pour pouvoir être géré par un hypertexte, un texte doit d'abord être transformé en un ensemble structuré d'informations. La base de données textuelles contient le texte, mais celui-ci a subi un premier traitement informatique : à chaque mot du texte correspond une référence interne, une « adresse logique » de l'ordinateur. La base de données textuelles est également structurée grâce à un certain nombre de délimiteurs de champs d'information. Citons, par exemple, les symboles qui déterminent le début et la fin d'un titre, le début et la fin d'un domaine d'application dans le cas d'une fiche terminologique, etc.

À ce premier niveau hiérarchique s'ajoutent d'autres niveaux, qui constituent eux-mêmes d'autres « nœuds », c'est-à-dire d'autres blocs d'information. Dans le cas de TACT, on distingue quatre autres niveaux : la liste complète des mots du texte, les critères de sélection, la liste des mots sélectionnés, les modes de disposition.

La liste complète des mots du texte constitue un bloc d'information. Elle ne demande aucune glose; ce sont tous les mots du texte. La liste des critères de sélection permet de spécifier selon quels critères les mots seront sélectionnés. TACT permet encore d'isoler automatiquement des suites de caractères en vue, par exemple, d'une étude des préfixes ou des suffixes, des mots de même famille, etc. TACT autorise également une sélection automatique de mots selon, par exemple, des critères de fréquence, selon aussi des critères de co-occurrence. Il est encore possible d'obtenir automatiquement à l'écran les passages d'un texte au sein desquels figurent, par exemple, les mots « amour » et « anarchie ». Tels sont quelques-uns des critères de sélection automatique de TACT. La liste des mots sélectionnés permet de choisir directement certains mots, de les pointer un à un dans la liste des mots du texte. Enfin, le dernier niveau est celui des modes de présentation des blocs d'information à l'écran.

Quatre modes de présentation de base sont disponibles : présentation des résultats sous forme de graphe de distribution, sous forme d'index, sous forme de concordance ou examen des mots dans leur contexte naturel, celui du texte de départ. Diverses options, correspondant elles aussi à des « nœuds », sont envisageables en tout temps : concordance de plusieurs lignes, index des mots avec contexte, index des mots avec fréquence ou sans fréquence, etc. L'impression des blocs d'information est évidemment possible et l'utilisateur peut résoudre le problème que pose parfois l'homographie.

Enfin, tous ces blocs d'information, accessibles en tout temps, rappelons-le, peuvent également être affichés dans des fenêtres contiguës. Toutes ces fenêtres, contiguës ou hiérarchisées, peuvent être déplacées, modifiées, rappelées à volonté.

En raison de cette souplesse, les hypertextes augmentent considérablement les possibilités des chercheurs en sciences humaines. Grâce à ces logiciels, la manipulation des données textuelles et l'étude des mots en contexte sont simplifiées. Le chercheur peut immédiatement vérifier ses intuitions; il peut passer de contextes très réduits à des contextes aussi larges que le texte lui-même. Enfin, si l'on songe aux possibilités

qu'offrent les lecteurs optiques, on peut envisager la constitution de bases de données textuelles de dimension réduite, de bases de données personnelles.

Dans le cas des logiciels conçus pour le traitement en lot, les résultats obtenus ne diffèrent guère de ceux obtenus par les 500 moines qui travaillaient sous la direction d'Hughes de Saint-Cher ou par les disciples de Calvin. L'apport de l'informatique demeure limité : l'ordinateur rend sans doute l'entreprise plus systématique et plus rapide, il en élimine surtout l'aspect fastidieux. Mais dans le cas des documents manipulés avec un hypertexte, le texte devient malléable. Aux yeux du chercheur, le texte acquiert, grâce à ce type de logiciels, une malléabilité que l'imprimé lui refusait.

25

Forme d'un dictionnaire électronique

Gaston GROSS

Laboratoire de linguistique informatique, Université Paris 13, Villetaneuse, France

Le Laboratoire de linguistique informatique travaille depuis quelques années à l'élaboration d'un dictionnaire électronique du français. Le travail a commencé avec le recensement et la description des noms composés. Nous avons ainsi fait une liste de plus de 100 000 noms composés correspondant à environ 500 types différents (ce travail a été fait par Michel Mathieu-Colas). Pour gérer cette masse d'information de façon unifiée, nous avons mis au point une série de champs constituant la description de chacune de nos entrées. Dans la mesure où nous constituons un dictionnaire électronique, nous ne pouvons avoir recours à une définition lexicographique pour rendre compte du « sens », c'est-à-dire de l'emploi des mots du lexique. Aucun logiciel n'est capable à l'heure actuelle de comprendre les définitions sémantiques qui figurent dans les dictionnaires destinés à la consultation humaine. Nous sommes donc contraints d'avoir recours à des informations formelles standardisées. L'objet de cette communication est de présenter la nature de ces informations. Chaque « article » du dictionnaire électronique est constitué de 13 champs : les six premiers sont de nature morphologique; trois autres décrivent les propriétés sémantiques et syntaxiques de l'entrée; les deux suivants définissent le domaine et le niveau de langue; les derniers enfin sont réservés à la traduction.

Les champs morphologiques

Le premier champ est constitué par l'entrée. Nous avons ensuite les champs suivants.

Le champ M représente le type morphologique. En particulier on note, pour les noms composés, le type auquel on a affaire : **NAdj, N de N, AdjN, Vn, V Dét N,**

etc. Rappelons que notre typologie comprend plus de 500 classes. Cette notation est importante pour la reconnaissance des composés dans les textes. En effet, pour un système informatique, un mot est défini comme une suite de caractères séparée par deux blancs. Cette définition exclut des séries discontinues comme les noms composés. Il est donc indispensable de noter les blancs qui ne sont pas des ruptures syntaxiques. Par la suite, nous indiquerons, soit ici soit dans un autre champ, si la suite est totalement figée, c'est-à-dire si elle n'a aucune variante et doit toujours être traduite de la même façon, ou s'il y des figements partiels (comme dans *élections présidentielles*, *élections cantonales*, *élections sénatoriales*).

Le champ G indique la catégorie grammaticale : s'il s'agit d'un nom, d'un adjectif, d'un verbe, etc. Pour les noms, nous signalons le genre et les indications de liberté ou non de nombre, ex : *nm*, *nf*, *nms*, *nmp*, *nfs*, *nfp*. Figure ainsi, pour des noms donnés, le caractère obligatoire du pluriel (*les eaux usées*, *les funérailles*) ou du singulier (*gentillesse*, *voûte céleste*).

Le champ F décrit les variations morphologiques des catégories susceptibles d'être fléchies : noms, adjectifs, verbes. Nous adoptons les types de flexions qui ont été élaborées au LADL (en particulier dans les dictionnaires *DELAS* et *DELAF*). Les noms et les adjectifs comprennent en français 80 types de flexions différentes (notés, par exemple, *N1*, *N2*, *N3*, *A1*, *A2*, *A3*) et les verbes 98 types (notés *V1*, *V2*, *V3*). Ces indications sont nécessaires pour la reconnaissance et la génération automatique de toutes les formes qui apparaissent dans les discours.

Le champ V rend compte des variantes graphiques. Il peut s'agir des variantes habituelles, comme les polygraphies (*événement* et *évènement*; *compte rendu* et *compte-rendu*; *steak* et *steack*) ou, dans le cas des noms composés de lectures non compositionnelles, comme dans *cordon bleu*, *livre blanc*, *pied de biche*, *va-et-vient*, etc.). On signalera si une suite figée appartient à un ensemble plus vaste, en fait s'il s'agit d'un sous-ensemble d'une expression (*la fleur de l'âge*, *à la fleur de l'âge*; *corps de garde*, *une plaisanterie de corps de garde*).

Les champs syntactico-sémantiques

Nous représentons ici l'ensemble des informations nécessaires à la reconnaissance et à la génération de phrases correctes. Notons que dans le cadre du lexique-grammaire, dans lequel nous nous plaçons, il n'y a pas de distinction de nature entre des « niveaux » syntaxique, sémantique et lexical. La notion fondamentale est celle d'« emploi » de l'opérateur qui regroupe ces trois niveaux. Dans cette théorie, l'opérateur et la phrase se confondent, puisque la phrase n'est que la réalisation des arguments possibles de l'opérateur. Les informations nécessaires à la description de chaque mot sont prises en charge par trois champs, respectivement les champs T, C, W.

Le champ T rend compte des propriétés des éléments lexicaux par des traits de nature syntactico-sémantiques comme :

- humains (hum)
- animal (an)

- inanimé concret (inc)
- inanimé abstrait (ina)
- locatif
- comptable ou non
- collectif (col)

On aurait ainsi :

hum : homme
hum, col : foule, jet-set, humanité, administration
an : chat
an, col : troupeau
in : graphique, plan, lettre
inc : caillou
ina : courage
in, col : tas, cuisine chinoise

Le champ C correspond aux classes d'objets que nous décrirons dans la deuxième partie de cet article. Il s'agit d'un concept nouveau de la grammaire désignant des ensembles sémantiques homogènes (*les vêtements*, *les moyens de transport*, *les outils*, *les sentiments*, par exemple) et qui ont des propriétés syntaxiques spécifiques. La mise en évidence des classes d'objets est très importante pour l'apprentissage des langues et pour la traduction.

Le champ W est réservé aux substantifs abstraits, c'est-à-dire prédicatifs. Comme les verbes, les substantifs prédicatifs sont des opérateurs. Ils ont donc, eux aussi, un domaine d'arguments qui leur est propre. En ce qui touche la logique des prédicats nous avons le schéma suivant :

opérateur (arg1, arg2, arg3)

Ce schéma est développé en phrase après deux opérations : la linéarisation qui place correctement les arguments de part et d'autre de l'opérateur et l'actualisation, qui, dans le cas des verbes, consiste à leur adjoindre un suffixe temporel. Cette possibilité n'existe pas pour les substantifs. En revanche, un substantif prédicatif est « actualisé » par des verbes à sens très « vagues », sémantiquement « vides » qui conjuguent pour ainsi dire ces substantifs. Il s'agit de verbes comme *faire*, *avoir*, *être*, *être Prép*, *donner*, *recevoir*, etc. appelés ***verbes supports*** :

Luc a décrit la ville.
Luc a fait la description de la ville.

Luc admire cette attitude.
Luc a de l'admiration pour cette attitude.

Luc voyage en Italie.
Luc est en voyage en Italie.

Luc a giflé Max.
Luc a donné une gifle à Max.

Luc a été giflé par Max.
Luc a reçu une gifle de Max.

L'introduction des verbes supports dans la description des langues est cruciale pour le traitement des langues et, en particulier, pour la traduction.

D'une part, les verbes supports permettent de rendre compte de la polysémie des opérateurs nominaux. Prenons le mot *charge*. Ses différents sens peuvent être mis en évidence par les différents supports que nous allons employer :

Luc a donné à Max la charge de faire ce travail.

Luc a la charge de faire ce travail.

Ce travail est une charge pour Luc.

Luc est en charge de ce travail.

Cette batterie est en charge.

Un notaire a une charge.

D'autre part, la construction à verbe support rend compte de la constitution des groupes nominaux et, en particulier, de la nature de la préposition :

Luc a du respect pour cette attitude.
Le respect de Luc pour cette attitude

Luc est en admiration devant cette attitude.
L'admiration de Luc devant cette attitude

Luc porte un jugement sévère sur Max.
Le jugement sévère de Luc sur Max

Luc a fait un salut à Max.
Le salut de Luc à Max

Observons au passage que dans la plupart des dictionnaires les arguments des prédicats nominaux sont donnés au génitif (subjectif ou objectif). À partir de ce génitif, on ne peut pas découvrir le verbe support qui a été effacé. L'inverse est vrai cependant. Il faut donc, pour définir une construction à opérateur nominal, donner systématiquement le verbe support adéquat.

Domaines et registres de langue

Nous décrivons ensuite des informations plus générales.

Le champ S rend compte du domaine dans lequel le mot en question prend sa signification. Pour le moment, nous nous servons des domaines que nous donnent les grands dictionnaires et les encyclopédies. Nous nous efforçons cependant de rendre les domaines les plus pointus possible.

Le champ R note les niveaux de langue comme *familier*, *populaire*, *littéraire*, *vieux*, *régional*, bref toutes les informations nécessaires à une bonne utilisation de l'élément lexical en question.

Les traductions

Les champs E et D reçoivent les traductions anglaises et allemandes, respectivement. Par la suite, une traduction dans d'autres langues est envisagée.

La notion de classes d'objets

Les relations sémantiques constituées dans le cadre de la phrase sont définies essentiellement par les compatibilités qui existent entre l'opérateur de cette phrase et son domaine d'arguments. Cette observation est mise clairement en évidence quand une forme morphologique donnée a plusieurs emplois. On voit alors que chaque emploi a des domaines d'arguments spécifiques. Le cas du verbe *regarder* constitue un exemple simple :

Les spectateurs regardent la rixe.
Cette affaire regarde les élèves.
La façade regarde la mer.
Je vais regarder ta dissertation.

Ces différents emplois (synonymes approximativement de *observer*, *concerner*, *donner sur*, *examiner*) ont tous des arguments différents. Ces arguments sont décrits, en particulier depuis le début de la grammaire générative, par des traits syntaxiques (ou plutôt syntactico-sémantiques) comme *humain*, *concret*, *abstrait*, *comptable*, *massif*, etc. Si l'on essaie de rendre compte ainsi d'un grand nombre d'opérateurs (des verbes, des substantifs ou des adjectifs prédicatifs, par exemple), le caractère restreint et peu discriminant dans leur combinatoire de ces traits a pour conséquence que des verbes seront décrits comme ayant un environnement identique alors que l'intuition nous pousse à les dissocier. Les verbes *manier* et *apprécier* auront la même distribution dans les phrases suivantes :

Luc a manié ce marteau.
Luc a apprécié ce marteau.

Il n'y a cependant aucune raison d'établir sur la base d'une distribution aussi restreinte une quelconque similitude linguistique. Les ouvrages qui font un classement de cette nature, comme en Allemagne les grammaires ou les dictionnaires de valence[1], donnent au lecteur une impression de répétition et de superficialité parce qu'ils ne rendent pas compte du sens précis de la relation entre l'opérateur et ses arguments.

D'autre part, le fait que l'on ait privilégié dans la tradition linguistique l'étude des opérateurs a eu pour corollaire une désaffection évidente pour les arguments et leur nature sémantique. Prenons l'exemple suivant :

Luc a regardé cet arbre.

1. Nous pensons en particulier aux travaux faits à Leipzig :
Wörterbuch zur Valenz und Distribution der Verben (HELBIG, SCHENKEL, VEB Bibliographisches Institut Leipzig, 1983); *Wörterbuch zur Valenz und Distribution der Substantive* (SOMMERFELD, SCHREIBER, VEB Bibliographisches Institut Leipzig, 1977). On peut renvoyer aussi à H. SCHUMACHER, 1986 : *Verben in Feldern*, De Gruyter, Berlin.

Le verbe *regarder* sera défini par le fait qu'il a comme sujet un substantif humain et pour objet un substantif concret, inanimé, comptable, etc. Les traits que nous venons d'indiquer serviraient aussi à rendre compte de l'emploi suivant :

Luc a planté cet arbre.

Or, il est évident que la relation entre *regarder* et le substantif *arbre* et celle entre *planter* et ce même substantif *arbre* ne sont pas du tout de même nature. Je sais beaucoup plus de choses sur la nature sémantique (ou pragmatique) du mot *arbre* à l'aide de l'opérateur *planter* qu'avec l'opérateur *regarder*. On pourrait objecter que le verbe *planter* est lui aussi ambigu du fait qu'il a dans son domaine d'arguments des substantifs de nature sémantique différente :

Luc a planté un clou dans le mur.
Luc a planté des choux dans son jardin.

À quoi on peut répondre que rien n'oblige à postuler qu'une classe sémantique naturelle soit définie par un seul opérateur. Si nous ajoutons à la définition de *arbre*, outre le verbe *planter* un opérateur comme *abattre* :

Luc a (planté + abattu) cet arbre

nous nous rendons compte que seuls les mots désignant des arbres sont susceptibles de figurer dans ce double environnement. On *plante un clou* ou *des choux* mais on ne les *abat* pas, on les *arrache* ou on les *récolte*. Nous venons de faire un trajet inverse à celui qui est fait habituellement. Au lieu de sous-catégoriser un opérateur à l'aide d'arguments, nous avons utilisé un opérateur (ou plutôt la combinaison de plusieurs opérateurs) pour définir un argument et, de façon plus générale, ce que nous appellerons une **classe d'objets**. Nous sommes amenés à considérer que les études de linguistique n'ont pas apporté beaucoup de lumière sur la relation sémantique ou pragmatique entre les opérateurs et leurs arguments. Il est assez clair que le verbe *regarder* n'a pas la même « proximité » avec le substantif *arbre* que le verbe *planter*. Le verbe *regarder* va sélectionner une classe énorme de substantifs qui n'ont en commun que le fait d'être des concrets, c'est-à-dire des objets susceptibles d'être vus. Ces objets n'ont entre eux aucune autre relation que celle-là mais diffèrent entre eux par un nombre indéterminé, mais très grand, de propriétés spécifiques. *Regarder* et *planter* sous-catégorisent donc de façon tout à fait différente leurs objets. Dans la mesure où, par rapport au substantif *arbre*, un verbe comme *planter* n'indique pas seulement qu'il s'agit d'un concret mais aussi à quel type de concret on a affaire, nous disons qu'il s'agit d'un **opérateur approprié**. La focalisation exclusive sur l'opérateur de la phrase et la description squelettique des arguments à l'aide de quelques traits réducteurs n'ont pas rendu possible une description approfondie du comportement linguistique des substantifs.

Prenons un autre exemple. Tous les dictionnaires, qu'ils soient scolaires, universitaires, destinés à un lecteur humain ou à un traitement automatique, utilisent le trait **humain** pour rendre compte du comportement linguistique des mots de cette catégorie. Mais y a-t-il quelque chose de commun entre des noms comme *homme, frère, menuisier, maire, régiment, parisien, salle, traître, tuberculeux, monsieur, accordéoniste, conducteur*, etc.? Tous ces substantifs désignant des humains constituent en fait des **classes d'objets** différentes. Par rapport à *homme* que l'on peut considérer comme un

humain « pur », *frère* est un relationnel, *menuisier* désigne une profession, *maire* une fonction, *régiment* est un collectif, *parisien* est un locatif, *salle* désigne un collectif par métonymie, *traître* est un nom de « qualité », *tuberculeux* désigne un malade, *monsieur* est un « appellatif », *accordéoniste* un instrumentiste, *conducteur* est un substantif prédicatif déverbal.

Ces classes d'objets ont chacune des propriétés syntaxiques particulières qui sont l'expression de chacune de ces classes d'objets.

Les noms de relations ont nécessairement un « complément », ce qui explique que leur détermination comprenne, en fait, un complément de nom ou un possessif :

*Luc est frère.
Luc est (mon frère + le frère de Paul).

Les noms de profession peuvent se passer de déterminant après le verbe *être* :

Luc est menuisier.

Ils ont aussi des « opérateurs appropriés » :

Luc (exerce + pratique) la menuiserie.
Luc est dans la menuiserie.

Ces mêmes opérateurs ne peuvent s'employer avec les substantifs désignant des fonctions, qui sont, elles, des compléments naturels de verbes comme *élire*, *désigner* ou *nommer*. On ne peut pas non plus, sauf ironie ou jeu de mots, leur adjoindre le complément *de profession* :

?Luc est maire de profession.

Les collectifs humains ont un comportement sémantique et pragmatique très particulier : ils sont en général compatibles avec des opérateurs désignant des activités intellectuelles ou psychologiques :

Le régiment est heureux de partir.
Le régiment a répondu positivement à cet appel.

Mais non avec ceux, par exemple, qui mettent en jeu des parties du corps :

Le professeur a mordu l'étudiant au mollet.
*Le professeur a mordu le régiment au mollet.

Les humains locatifs sont associés à des toponymes et ont une syntaxe particulière :

Luc est parisien.
Luc (est de + habite + vient de) Paris.

Les métonymies à partir de locatifs ont en partie les mêmes propriétés négatives : les propriétés spécifiques des noms partie-du-corps sont suspendues, il y des restrictions sur l'emploi du pluriel, etc.

Les substantifs de « qualité » peuvent fonctionner comme attribut (Luc est un traître), la forme emphatique est possible (C'est un traître), ce sont des « vocatifs » (Espèce de traître ! Traître !), ce sont des compléments naturels de verbes comme *traiter N1 de N2*.

Les noms désignant des malades ont aussi un comportement linguistique autonome, comme on le voit dans ces paraphrases :

Luc est tuberculeux.
Luc a la tuberculose.
Luc (souffre + est mort) de tuberculose.

Les « appellatifs » sont des vocatifs par excellence. C'est ce qui fait que leur utilisation en position d'argument puisse être considérée comme déviant de la norme, comme dans le langage enfantin :

« Va dire bonjour au Monsieur! »

Une classe comme celle des instrumentistes est bien définie du point de vue syntaxique, sémantique et pragmatique :

Un accordéoniste joue de l'accordéon.

Il reste cependant une ambiguïté : un instrumentiste peut être soit un amateur soit un professionnel. Dans ce dernier cas, les ensembles ne sont pas disjoints. Enfin, les substantifs prédicatifs forment une immense classe qui a la même extension que les verbes qui leur sont associés.

On voit donc que la classe des humains regroupe des sous-ensembles qui ont des particularités linguistiques bien définies. Ces sous-ensembles constituent des classes sémantiques et pragmatiques qui les motivent et que nous appelons classes d'objets. Nous pensons que ce sont ces classes qui sont le niveau d'analyse adéquat et non pas un trait aussi général que *humain* qui banaliserait toutes ces particularités et ne rendrait pas compte des faits de langage que nous venons d'évoquer rapidement.

Sur le modèle de ce que nous venons de dire, nous pouvons ainsi mettre en évidence des classes d'objets, qu'ils soient abstraits ou concrets. Nous insistons cependant sur le fait que ces classes ne sont pas décrites indépendamment de leur comportement réel dans le discours, c'est-à-dire indépendamment de la syntaxe. Les classes d'objets relèvent dans notre esprit à la fois de la syntaxe, de la sémantique et de la pragmatique, c'est-à-dire en fin de compte de la cognition.

Les classes d'objets ont un assez grand pouvoir explicatif sur la façon dont les objets sont perçus par l'esprit humain dans le cadre de la langue. Soit la classe des **habits**. Les opérateurs appropriés sont les suivants :

Luc a (mis + enlevé) sa veste.
Luc a changé de veste.
Luc porte une veste.

Sur la base de ces quatre verbes, il est possible de dresser la liste de tous les habits du français. Une recherche méthodique, mais non exhaustive, nous en a fait trouver environ 800. La définition que nous venons de donner inclut cependant un mot comme *lunettes*. Pour la langue, les *lunettes* sont un vêtement :

Luc a (mis + enlevé) ses lunettes.
Luc a changé de lunettes.
Luc porte des lunettes.

Cette assimilation opérée par la langue permet peut-être de rendre compte d'un emploi tout à fait inattendu :

Luc a chaussé ses lunettes.

Applications à la traduction

Si l'on reprend l'opposition que nous avons établie entre les opérateurs « généraux » comme *regarder*, qui ne découpent parmi les substantifs que des ensembles très vastes, peu spécifiques et correspondant à des traits syntactico-sémantiques comme *humain* et les opérateurs « appropriés » qui constituent des classes d'objets, on se rend compte que cette opposition a une importance inattendue dans le domaine de la traduction.

En effet, le premier type de verbe ne pose pas de problèmes particuliers de traduction. Une phrase comme :

Les enfants ont regardé le spectacle

pourra se traduire terme à terme et le verbe aura le même sens quel que soit le complément, à condition qu'il puisse être interprété comme un objet concret ou un événement. Tout se passe comme si les opérateurs à très large spectre correspondaient, dans les langues, à une expérience commune de l'humanité et que, de ce fait, la relation entre opérateurs et arguments était, pour ainsi dire, stable et commune aux langues.

En revanche, les opérateurs appropriés posent des problèmes délicats de traduction : leur comportement n'est pas toujours prédictible. De ce point de vue, ils sont l'objet de difficultés de traitement similaires aux mots composés, dont le sens n'est pas non plus prédictible. Prenons la classe d'objets <*vêtement*>. Parmi les verbes appropriés nous avons relevé à l'instant *porter* :

Luc porte un chapeau neuf

qui ne se traduit pas par *carry* mais par un verbe spécifique *wear* :

Luc is wearing a new hat.

On observera qu'en anglais aussi les *lunettes* font partie du *vêtement* puisqu'ils ont le même opérateur approprié :

Luc wears spectacles.

Le verbe *porter* s'il s'applique à la classe d'objets <*blessure*> aura une traduction différente :

Luc porte des cicatrices.
Luc bears scars.

Si la classe d'objets est une <*partie du corps*> (ici spécifique), la traduction sera encore différente :

Le chameau porte deux bosses.
The camel has two humps.

Prenons un exemple qui met en jeu l'apprentissage du français. Les verbes appropriés de la classe d'objets <*moyen de transport*> sont, entre autres, *voyager par, prendre* :

Luc a voyagé par le train.
Luc a pris le train.

Dans ce cas, le verbe *prendre* a comme synonyme *emprunter* :

Luc a emprunté le train.

Cette synonymie n'est cependant possible qu'avec les moyens de transport collectifs. En effet une phrase comme :

Luc a emprunté le vélo

signifie qu'il l'a demandé à un voisin ou qu'il l'a volé. La notion de classes d'objets a donc un intérêt évident dans l'enseignement.

Voici d'autres exemples d'utilisation des classes d'objets pour désambiguïser des opérateurs :

Opérateurs verbaux :

prendre (objet) : *etwas ergreifen*

prendre <photo> : *ein Foto machen*
prendre <moyen de transp> : *ein Taxi, Flugzeug nehmen*
prendre <voie> : *Wir sind (auf der) Autobahn gefahren*
prendre <coups> : *Schläge kriegen*
prendre <ordre> : *ein Befehl übernehmen*

tirer (voiture) : *einen Karren ziehen*

tirer <gibier> : *schießen*
tirer <imprimé> : *drucken*
tirer <temps> : *abreißen (drei Jahre Gefängnis abreißen)*
tirer sur le <couleur> : *ins < > spielen*

poser (un objet) : *etwas hinlegen*

poser <avion> : *ein Flugzeug aufsetzen*
poser <revêtement de sol> : *Fliesen legen*
poser <rideau> : *eine Gardine aufhängen*

avoir (objet) : *besitzen, haben*

avoir (abstrait) : *Mut haben*
avoir <action> : *einen Kuß bekommen*
avoir <prix, récompense> : *den ersten Preis erhalten*
avoir <moyen de transport> : *den Zug gerade bekommen*
avoir <gibier> : *treffen (fusil)*
avoir <âge> : *X Jahre alt sein*

faire (objet) : *etwas herstellen*

faire <Poss habitation> : *putzen*
faire <instrument de mus> : *Klavier spielen*

faire <sport> : *Sport treiben, Fußball spielen*
faire <langue> : *Englisch studieren, lernen*
faire <matière enseignée> : *Medizin studieren*
faire <monument> : *ein Schloß besichtigen*
faire <maladie> : *Fieber haben*
faire <culture> : *anbauen : Er baut nur Weizen an*
faire <métier> : *ein guter Lehrer (werden, sein)*
faire <prix> : *kosten*
faire <contenu> : *dieser Tank faßt fünfzig Liter*
faire <poids> : *haben, schwer sein*

pondre (œuf) : *legen*

pondre <enfants, plo> : *x Kinder kriegen*
pondre <écrit> : *verfassen, zusammenschmieren*

pousser (personne, objet) : *(an)stoßen*

pousser <voiture> : *schieben*
pousser <cri, parole> : *ausstoßen*
pousser <porte> : *aufstoßen*
pousser <moteur> : *hochjagen*

lancer (objet) : *werfen*

lancer <bombe> : *abwerfen*
lancer <satellite> : *starten*
lancer <cri> : *ausstoßen*
lancer <artiste> : *fördern*
lancer <produit> : *einführen*
lancer <gibier> : *aufjagen*

défricher <forêt> : *einen Wald roden, urbar machen*
défricher <sujet> : *aufbereiten, in Angriff nehmen*

battre (humain) : *prügeln*

battre <adversaire> : *schlagen*
battre <tapis> : *einen Teppich klopfen*
battre <vêtement> : *ein Kleid ausklopfen*
battre <plat> : *Teig schlagen*
battre <céréales> : *Korn dreschen*

piloter <avion> : *fliegen*
piloter <moyen de transp> : *fahren*
piloter <bateau> : *lotsen*

charger (véhicule) : *beladen*

charger <batterie> : *laden*
charger <four> : *einen Hochofen begichten*
charger <ennemi> : *angreifen*

essuyer <vaisselle> : *das Geschirr abtrocken*
essuyer <superficie> : *abwischen*
essuyer <meuble> : *abstauben*
essuyer <lunettes> : *putzen*
essuyer <événement malheureux> : *mitmachen*
essuyer <attaque> : *Verluste erleiden*
essuyer <partie du corps> : *sich X abwischen*

égrener <céréales> : *entkörnen*
égrener <fruit> : *rebeln*
égrener <chapelet> : *den Rosenkranz abbeten*

emprunter (argent) : *sich Geld leihen*

emprunter <outil> : *ich nehem kurz ihren Hammer*
emprunter <mot> : *ein Wort (aus N) entlehen*
emprunter <voie> : *eine Straße benutzen, befahren*
emprunter <moyen de transp> : *nehmen*

jouer à (jeu) : *spielen*

jouer <sentiment> : *Verzweiflung mimen*
jouer <air> : *spielen*
jouer <instrument> : *Geige spielen*
jouer <spectacle> : *Rolle aufführen*
jouer les <rôles> : *die Naive spielen*
jouer à <sport> : *Fußball spielen*

Opérateurs adjectivaux :

abandonné <animal> : herrenlos
abandonné <lieu> : *verlassen*
abandonné <chantier> : *aufgelassen*
abandonné <attitude> : *gelöst*

abondant <moisson> : *reich*
abondant <nourriture> : *reichlich*
abondant <pluie> : *ausgiebig*
abondant <langue> : *reich*
accablant <chaleur> : *drückend*
accablant <douleur> : *quälend*

accessible <personne> : *empfänglich*
accessible <œuvre> : *verständlich*
accessible <lieu> : *erreichbar*

actif <personne> : *aktiv*
actif <vie> : *betriebsam*
actif <militaire> : *aktiv*
actif <médicament> : *wirksam*

actif <verbe> : *im Aktiv stehend*
actif <travailleur> : *berufstätig*

aigre <fruit> : *sauer*
aigre <alcool> : *sauerlich*
aigre <vent> : *scharf*
aigre <voix> : *grell*

aigu (objet) : *spitz*

aigu <pointe> : *spitz*
aigu <voix> : *grell, schrill*
aigu <note> : *hoch*
aigu <douleur, sensation> : *heftig, stark*
aigu <maladie> : *akut*
aigu <conflit> : *heftig*
aigu <intelligence> : *scharf*

anémique <malade> : *blutarm*
anémique <plante> : *kümmerlich*
anémique <style> : *farblos*

Applications à la traduction automatique : remplacement de l'arbre sémantique

Il existe au moins un consensus parmi les linguistes qui s'occupent de près ou de loin de traduction automatique : sans composante sémantique, la traduction automatique ne peut pas marcher. Les informations sémantiques y sont prises en charge par des codes sémantiques (800 environ pour un système de traduction automatique comme SYSTRAN, taxonomisés en six arborescences : *Choses*, *Actions*, *États*, *Qualités*, *Lieux*, *Positions relatives*). Le monde est dans ce système divisé en six grandes sphères et chaque terme-mot doit appartenir à l'une de ces sphères. À l'intérieur de chaque taxonomie, à chaque division correspondent un « nœud » et un code qui ne couvrent qu'un seul concept. Si, par exemple, un terme reçoit un code X, il recevra automatiquement les codes afférents aux nœuds supérieurs jusqu'au taxon de tête. Si un seul de ces codes ne s'applique pas au terme en question, le code inférieur, qui semblait pourtant convenir, ne doit pas lui être appliqué, à moins que l'on utilise un « stoppeur », suivi d'un code issu d'un des cinq autres taxons applicables.

Par exemple, le code <PROF> (profession) fait partie de l'arbre des <Qualités>, mais on peut l'appliquer à des personnes humaines, à condition de le bloquer par <CHAR> et de lui donner le code <HUMANS>, de l'arbre des <Choses>. Prenons d'autres exemples. RÉSOLVABLE recevrait selon le système sémantique SYSTRAN, le code <SOLUT> (solution) et les nœuds supérieurs <RESMENT> (result of a mental action), <RESPON> (response) et <QUAL>, le taxon, alors que SOLUTION recevrait les codes <SOLUT>, <CHAR> (le stoppeur), <PROCESS>. Mais ne devrait-il pas recevoir simplement <RESAC> ou <CHGST> (change of state) du

taxon <PROCESS>? À moins que ce ne soit un <CHCOMP> (chemical compound) dans l'arbre de <THINGS> ou un <COMP> (composition) tout court! Que choisir parmi les codes

<DIRMOT> (directional motion)
<DMFR> (directional motion from; leaving)
<DMTFD> (directional from/to; definite)
<DMTFI> (directional from/to; non specific)
<DMTO> (directional motion to a place)

quand il s'agit de coder les verbes *voyager* ou *rouler*? On imagine les affres pour le codeur!

On comprend bien vite que cet extrême raffinement de concepts hiérarchisés ait donné lieu, dans la réalité, à une utilisation sommaire de moins de 10 pour cent des codes, ceux dont le concept ou parfois seulement le symbole mnémonique semblait clair et bien délimité. La pratique réelle du codage a évolué vers une petite cinquantaine de codes évidents comme :

NATION : nationalité
CHELM : élément chimique
PROF : profession
PLANT : végétal
ANIMAL : animal
HUMANS : être humain
MOTION : mouvement
etc.

On s'aperçoit qu'il s'agit là beaucoup plus de classes d'objets, au sens où nous les avons définies plus haut, que de concepts hiérarchisés. Et s'il est important de savoir que attraper un <ANIMAL> n'est pas attraper une <MALADIE>, faut-il vraiment savoir que <ANIMAL> est un hyponyme de <INDIVIDU> (opposé à <GROUPE>), lui-même hyponyme de <ANIMÉ> (opposé à <INANIMÉ>), de <PHÉNOMÈNE> (opposé à <CONCEPT>) et de <CHOSE> (opposé aux autres taxons comme <ACTION> ou <QUALITÉ>)? Alors que <MALADIE> serait un hyponyme de <ÉVÉNEMENT> (opposé à <SUBSTANCE PHYSIQUE>), de <INANIMÉ> et de <PHÉNOMÈNE>, autrement dit, finissant en fin de compte dans la même vaste catégorie que <ANIMAL>.

Ces réflexions et ces exemples nous amènent à constater ce que les arbres sémantiques ont d'abstrait, d'artificiel et d'inutile, c'est-à-dire de redondant quand il s'agit d'obtenir une traduction adéquate pour un domaine d'arguments particulier d'un opérateur.

Enfin, les taxonomies supposent que l'on trouve un code adéquat pour un terme dans un des arbres sémantiques. Les mélanges sont fortement déconseillés. Or beaucoup de mots sont à cheval sur plusieurs concepts ou, plus exactement, sur plusieurs classes d'objets et, tout en étant les arguments possibles de certains prédicats, influent sur des structures de natures différentes. Une *cotte de maille* est à la fois un <objet métallique> et un <vêtement>, puisqu'une cotte de maille peut *rouiller* et qu'on peut *porter* ou *mettre une cotte de maille*.

Les classes d'objets permettront de décrire les substantifs avec beaucoup plus de souplesse : l'objectif est d'établir une bijection entre chaque opérateur et son domaine d'arguments, décrits selon les classes d'objets. On sait d'expérience qu'un assez grand nombre de verbes de fréquence élevée posent des problèmes terribles de traduction ou de traitement automatique (de même d'ailleurs que d'apprentissage, en cas de langue seconde, comme nous l'avons vu plus haut). Ces problèmes peuvent être résolus si ces verbes sont décrits à l'aide des classes d'objets. Ainsi pour *prendre* :

prendre <aliment, boisson> : prendre un steak, du thé
prendre <moyen de transport> : prendre le train
prendre <contenant> de <aliment, boisson> : prendre un verre
prendre <voie> : prendre l'autoroute, un sentier

De même peut-on rendre compte de *être en* :

être en <vêtement> : être en pyjama
être en <matière> : être en fer, bois, verre
être en <moyen de transport> : être en train
être en <matière enseignée> : être en Lettres, Droit
être en <pays> : être en France
être en <locatif> : être en prison

Les classes d'objets permettent de monosémiser des verbes qui ont des emplois particulièrement nombreux.

Dans ces conditions, le codage ayant été fait à la fois sur l'opérateur et sur chaque argument, la compatibilité sémantique sera respectée. Il nous faudra donc faire la liste la plus complète possible de chaque classe d'objets. Ce travail est nécessairement de longue haleine. En effet, il convient, pour chaque classe d'objets, d'examiner méthodiquement tout le lexique des noms. On ne voit pas comment on pourrait établir ces classes autrement. Il n'y a, à l'heure actuelle, aucun travail, à notre connaissance, qui soit en chantier. Ce travail de recension ne peut pas être informatisé. Seul un cerveau humain est en mesure de déterminer si un élément fait partie ou non d'une classe d'objets déterminée.

Domaines d'applications de la notion de classes d'objets

Le premier intérêt de ce projet est de nature théorique. La mise au point de la notion de classes d'objets, définies de façon interne au système de la langue et à l'aide d'éléments lexicaux et non de traits abstraits, permet de poser de façon nouvelle la relation entre les différents niveaux d'analyse linguistique : la syntaxe, le lexique, la sémantique et la pragmatique. Pour la description des éléments linguistiques, cela reviendra à ouvrir un nouveau champ dans les banques de données. À côté de ceux rendant compte des traits de sous-catégorisation des noms (humain, concret, abstrait, etc.) et des domaines (chimie, biologie, médecine, psychologie, psychiatrie, etc.), nous proposons d'ajouter le champ des classes d'objets. Le premier travail que nous venons de faire sur un millier de mots a montré que la combinaison de ces trois champs a permis d'éviter pour la description de ces mots toute ambiguïté. Du point de vue

conceptuel, nous posons les relations sémantiques entre opérateurs et arguments de façon nouvelle.

Sur le plan des réalisations, on peut envisager les applications suivantes :

a) Dans le domaine de l'enseignement des langues première et surtout seconde, qu'il s'agisse de télé-enseignement ou de l'enseignement assisté par ordinateur (EAO), la description des classes d'objets permet de rendre compte de façon claire de constructions considérées comme bizarres ou idiosyncratiques (exemple : *chausser ses lunettes*).

b) Dans le domaine de la traduction automatique, les contacts que nous avons eus avec la société SYSTRAN ont montré que le remplacement de l'arbre sémantique par les classes d'objets permettrait d'éviter un très grand nombre de fautes dues à la non-reconnaissance du sens du verbe. Les adaptations ne sont pas d'une grande complexité. Il faudrait :
   - créer des tables de conversion entre nouveaux codes et anciens codes;
   - supprimer les nœuds supérieurs;
   - créer des routines de transfert sur la base de codes de classes d'objets.

c) Enfin, on peut envisager sérieusement de créer un « documentateur automatique ». On a constaté depuis le traitement automatique des langues que l'essentiel de l'information d'un texte est pris en charge par les substantifs. Nous envisageons de coder tous les noms (environ 50 000 noms simples et plus de 100 000 noms composés que nous avons recensés par ailleurs) au regard des trois champs dont nous avons parlé :
   - les traits syntactico-sémantiques;
   - les domaines;
   - les classes d'objets.

Une fois ce dictionnaire électronique constitué, on peut envisager de faire passer un texte donné (un article de journal ou un texte scientifique, par exemple) dans ce dictionnaire. Tous les substantifs recevraient ainsi les marques correspondant à chacun de ces trois champs. En cas d'ambiguïté le codage des mots du contexte permettra de choisir la solution adéquate. Comme on sera en mesure de connaître le domaine et la classe d'objets de tous les noms du texte, on pourra ainsi prendre connaissance du contenu de ce texte. Si l'on associe à ce traitement des procédures statistiques sur la fréquence des domaines et des classes, on pourra savoir si un texte parle de chirurgie mais aussi s'il s'agit d'une présentation d'instruments chirurgicaux (classe d'objets) ou d'actes chirurgicaux (autre classe d'objets).

Conclusion

Si l'informatique offre, à l'avenir, pour le traducteur une possibilité d'avoir un accès aux données beaucoup plus rapide qu'auparavant, il ne faut pas oublier qu'une traduction est un passage d'une langue naturelle à une autre langue naturelle. Pendant longtemps, le traitement automatique du langage, en particulier la traduction auto-

matique à ses débuts, était censé pouvoir résoudre les problèmes de façon quasi indépendante de la nature de la grammaire. Nous savons maintenant que nous ne pouvons pas nous passer de dictionnaires. Les expériences de traduction automatique en grandeur réelle (style SYSTRAN) montrent de façon évidente que la plupart des fautes sont liées au lexique. Si l'on décrit les mots, dans un cadre théorique comme celui du lexique-grammaire qui ne conçoit pas la grammaire, la sémantique et le lexique à des niveaux différents, on décrit alors les opérateurs en fonction d'emplois, or la matière première du traducteur, ce n'est pas les mots, ce sont les emplois.

Bibliographie

CHOMSKY N. (1970) « Remarks on Normalisation », *Readings in English Transformational Grammar*, Jacobs Rosenbaum (éd.), Toronto, Waltham.

DUBOIS J. (1961) « La notion d'unité sémantique complexe et sa neutralisation dans le lexique », *Cahiers de Lexicologie II*, Paris, Didier.

FRASER B. (1970) « Idioms within a Transformational Grammar », *Foundations of Language*, VI.

MAKKAI A. (1978) « Word Structure », *Universal of Human Language*, J.H. Greenberg (éd.), Stanford, Stanford University Press.

GROSS G. (1986) « Typologie des noms composés », *Rapport ATP*, CNRS, Université Paris 13.

GROSS G., VIVES R. (1986) « Les constructions nominales et l'élaboration d'un lexique-grammaire », *Langue Française*, 69, Paris, Larousse.

GROSS G. (1988) « Degré de figement des noms composés », *Langages*, n° 90, Paris, Larousse.

GROSS G. (1989) *Les constructions converses du français*, Genève, Droz.

GROSS M. (1981) « Les bases empiriques de la notion de prédicat sémantique », *Langages* 3, Paris, Larousse.

GROSS M. (1982) « Une classification des phrases figées du français », *Revue québécoise de linguistique*, 11, 2.

GROSS M. (1990) *Grammaire transformationnelle du français : syntaxe de l'adverbe*, Paris, Asstril.

HARRIS S.Z. (1976) *Notes du cours de syntaxe*, Paris, Le Seuil.

LEVI J.N. (1973) « Where Do all those Other Adjectives Come from? », *Papers from the Ninth Regional Meeting : Chicago Linguistic Society*, Chicago.

LEVI J.N. (1978) *The Syntax and Semantics of Complex Nominals*, New York, Academic Press.

ROHRER CH. (1967) *Die Wortzusammensetzung im modernen Französisch*, Diss., Tübingen.

RUWET N. (1972) *Théorie syntaxique et syntaxe du français*, Paris, Le Seuil.

WARREN B. (1978) *Semantic Patterns of the Noun-Noun Compounds*, Gothenburg, Gothenburg Studies in English 41.

WARREN B. (1984) *Classifying Adjectives*, Gothenburg Studies in English 56, AWE International.

26

À la recherche d'écosystèmes terminologiques

Daniel BLAMPAIN, Philippe PETRUSSA, Marc VAN CAMPENHOUDT

Institut supérieur des traducteurs et interprètes de la Communauté française de Belgique à Bruxelles

L'intitulé de cette communication est né du thème de ce colloque consacré à « l'environnement traductionnel ». Il correspond bien à la démarche que notre équipe mène dans un milieu universitaire de formation et de recherche.

Notre intention est de présenter les objectifs et les premiers résultats d'une recherche qui se déploie à partir d'un département de linguistique appliquée, en collaboration avec les départements d'informatique, de langues et divers milieux scientifiques et techniques. La pertinence de l'analogie posée par le titre apparaîtra progressivement.

Nous sommes partis d'un constat sur la place occupée par la terminologie dans l'environnement traductionnel, qu'il s'agisse du milieu professionnel ou d'un milieu de formation : la LUXURIANCE de la terminologie est telle que l'utilisateur a peine à s'y retrouver.

1. Le développement des sciences et des techniques a multiplié les besoins en terminologie et les produits foisonnent « sauvagement », des entreprises aux institutions.
2. Le mot « terminologie » est devenu tellement polysémique que l'on oscille chez les praticiens entre la tentation de ramener la terminologie à une simple question de lexique, voire de lexicographie, et celle de l'assimiler à une science du langage, à une réflexion théorique dont ils n'ont que faire dans la pratique de l'urgence qui est la leur.
3. Depuis près de vingt ans, des banques de terminologie se déclarent au service du traducteur. Si nous laissons de côté les produits imprimés, nous pouvons

dire que dans un bureau de traduction bien équipé ou dans un institut comme le nôtre, la consultation des banques se résume, pour des raisons économiques, à celle de TERMIUM sur CD-ROM et d'EURODICAUTOM. Mais même au sein de ces deux banques, le concept de « luxuriance » est tellement d'application que le traducteur retrouve parfois la solitude et le découragement de l'explorateur au sein des forêts vierges.

3.1. Ces banques sont encombrées. Elles ont accumulé des termes de langue courante et de langue de spécialité. L'obsolescence ternit l'intérêt de beaucoup d'interrogations.

3.2. La sélection des domaines reste générale et peu efficace.

3.3. Elles sont construites à partir de systèmes informatisés anciens et lourds, voire peu rentables. La sélection de fiches ou de données à l'intérieur de fiches reste un problème qui témoigne de l'archaïsme du système de consultation.

L'utilité de ces banques — nous le voyons tous les jours — est pourtant réelle. Elles sont le fruit d'une longue expérience de traduction qui sert nos étudiants, mais pour ces derniers l'urgence et la rentabilité ne se posent pas dans les mêmes termes que pour le traducteur professionnel. La consultation des sources, fondamentale pour un questionnement scientifique ou simplement pour une traduction sérieuse, correspond par exemple dans TERMIUM à une manœuvre qui doit se faire crayon en main et qui prend beaucoup trop de temps. Que penser d'autre part d'un traitement dit « terminologique » qui présente le mot (le terme?) accompagné d'une définition, avec ou sans référence, suivi de ses équivalents dans d'autres langues avec la traduction de ladite définition?

4. Enfin, le concept de « terminologie », déjà fragmenté entre des méthodologies différentes, s'est trouvé plongé dans la luxuriance du champ des industries de la langue, où l'on entend chanter les sirènes de la TAO, chant programmé par des « logiciels toujours plus performants » qui alimenteront la « station de travail du traducteur de l'an 2001 ». « Le miroir aux alouettes », s'exclamait un traducteur professionnel au colloque de Paris consacré à la terminologie et aux industries de la langue (19-20 janvier 1989). Il reposait en termes crus — avec une audace incroyable dans ce contexte — cette question : « Combien de traducteurs savent ce qu'est la terminologie et savent *en faire*? »

Devant cette luxuriance, qui ne peut apparaître comme un signe de bonne santé qu'aux esprits romantiques, à l'écart des mythes, renversant humblement la TAO en OAT (outils d'aide à la traduction), nous proposons de saisir la problématique de la terminologie à un niveau précis et expérimental, avec un certain nombre d'exigences scientifiques auxquelles les « entreprises » de terminologie sont aujourd'hui dans l'impossibilité de répondre. Seul le milieu universitaire peut prendre en considération la terminologie en tant qu'activité linguistique descriptive, non soumise à des projets politiques ou normatifs, et contribuer à la mise au point d'« écosystèmes terminologiques ».

Par rapport à la « luxuriance » décrite ci-dessus, nous retrouvons donc la pertinence du concept d'« écosystème », qui implique non seulement l'étude d'équilibres et la

partition d'une entité en systèmes relativement autonomes dont il est possible d'analyser la structure et le fonctionnement, mais aussi la mise en évidence de mécanismes de régulation qui permettent une adaptation optimale aux conditions changeantes de l'environnement.

En d'autres termes, le traducteur confronté aux problèmes posés par les langues de spécialité et amené à devoir maîtriser — plus rapidement que par de simples lectures de familiarisation — les notions fondamentales de domaines et de sous-domaines, devrait avoir à sa disposition des instruments terminologiques informatisés, susceptibles d'être constamment remis à jour, qui lui permettent d'accéder aux réseaux notionnels structurant ces micro-domaines de haute spécialité.

Les options théoriques sur lesquelles reposent notre travail sont les suivantes :

1. Toute démarche terminologique doit se fonder réellement sur les notions. La terminologie sera comprise comme la totalité structurée des notions d'un *domaine de spécialité*. La plate-forme théorique mise au point par Wüster et reprise par Felber se retrouve ici. Les conséquences du choix de cette approche conceptuelle, que l'on trouve aussi bien en sémantique ou en psycholinguistique, sont importantes :

   1.1. La terminologie ne peut être assimilée à une lexicographie de domaines spécialisés.

   1.2. Elle ne se ramène pas à la rédaction de fiches, aussi complètes (ou utopiques) soient-elles, ou à l'élaboration d'arborescences documentaires.

   1.3. Elle ne doit pas se limiter à la problématique posée par Wüster que l'on trouve aujourd'hui fétichisée, peut-être parce que la terminologie en tant que science du langage n'a guère qu'un demi-siècle d'existence, comme le rappelait A. Clas à Genève, lors du colloque sur « L'enseignement de la terminologie » (septembre 1988).

2. La fonction cognitive et classificatrice de la terminologie est privilégiée. La terminologie exprime un système de connaissances. Les notions sont à percevoir dans le sens épistémologique d'ensembles de caractères qui nous permettent de reconnaître des objets et dans le sens logique d'éléments d'ordonnancement du savoir.

   En somme, la terminologie conçue par référence à de véritables systèmes notionnels reconnus dans des domaines thématiquement homogènes constitue le seul moyen de revitaliser le mot « TERME », dont la valeur sémantique est conditionnée par sa place dans un système.

3. À une représentation adéquate de la connaissance transmise par les langues de spécialité doivent être alliés les progrès réalisés en matière de logiciels.

   A. Clas, dans un article publié dans *Meta* (32-2) en 1981 (« Les nouveaux lexiques ou une stratégie de création de mini-banques »), déclarait, lui qui a porté TERMIUM sur les fonts baptismaux, qu'il croyait en l'avenir de mini-banques sur micro-ordinateur qui ne visaient plus la collectivité mais l'individu et qui répondaient au besoin de trouver équivalents, synonymes, abréviations...

Nous avons le souci d'aller au-delà de cette demande. Il est d'ailleurs significatif que notre démarche n'est pas isolée et que le Secrétariat d'État d'Ottawa s'intéresse aujourd'hui, après vingt ans de pratique terminologique, à un logiciel qui se réfère aux réseaux notionnels (CODE, projet pilote établi en collaboration avec l'Université d'Ottawa).

Le logiciel *Termisti* (version expérimentale)[1]

Le logiciel *Termisti*, conçu dans le cadre d'une recherche, n'est pas le énième produit propulsé sur le marché des gestionnaires de données terminologiques. Afin de mieux comprendre la structure notionnelle des vocabulaires de spécialité, il se fonde sur des acquis théoriques et informatiques stables et décrit des micro-glossaires homogènes. Cette pratique scientifique repose sur divers principes :

1. Tous les idiomes envisagés doivent pouvoir servir de langue source ou de langue cible dans le cadre de n'importe quelle combinaison (3 langues = 6 sens de traductions).
2. Ceci implique une approche notionnelle où la notion est conçue comme « internationalement conditionnée » : c'est la confrontation des notions présentes dans chaque langue envisagée qui détermine le nombre de notions à prévoir dans la banque pour rendre compte du réel. En d'autres termes, il y aura autant de notions homonymes en langue source qu'il y a de notions correspondantes en langue cible.
3. La fiche terminologique doit distinguer le savoir consacré à la notion et le savoir consacré aux diverses désignations de cette notion dans les différentes langues envisagées.
4. Au sein d'un micro-glossaire, les notions peuvent être reliées par des liens verticaux et horizontaux formant une arborescence rigoureusement structurée. Ce mode d'organisation notionnelle constitue une aide à la consultation pour le traducteur et permet de rendre compte de notions hyponymes rarement définies dans les sources spécialisées du fait qu'elles sont situées au bas d'une hiérarchie espèce-genre.
5. Des règles de fonctionnement des arborescences peuvent être exploitées dans une perspective de gestion automatisée du savoir. En effet, l'organisation notionnelle du savoir n'est qu'une étape préalable au travail du traducteur. Le rôle du logiciel est de proposer une démarche intelligente d'exploitation du contenu de l'arborescence. Ainsi, il pourrait proposer des modifications de liens lors de l'ajout d'une nouvelle langue (c'est-à-dire de nouvelles notions) ou résoudre les impossibilités de traduction d'une notion en exploitant le réseau. Le projet tente actuellement de découvrir les algorithmes qui, dans l'avenir, permettront d'accomplir de telles tâches.

1. La version expérimentale de *Termisti* a fait l'objet d'une démonstration au cours de l'exposé.

6. Le savoir notionnel engrangé peut être utilisé dans de nombreuses directions, notamment pour la production de glossaires et de lexiques modulables par le choix des champs et des relations en fonction des besoins du traducteur. On distinguera donc l'engrangement du savoir terminologique et son exploitation strictement terminographique.

Répondre à ces exigences, c'est réfléchir à des règles pertinentes qui se dégageront de l'observation de terminologies dans leur usage réel. Les contextes et les définitions véhiculent, en ce sens, de nombreuses informations sur la place de la notion dans l'arborescence.

Contenu

Fiche

La fiche de *Termisti* applique les principes notionnels et tente de développer les acquis théoriques de l'école de Vienne. Une même notion peut donc être désignée dans une infinité de langues et une même langue présenter un nombre indéterminé de termes pour désigner ladite notion.

La plupart des champs correspondent à ceux qui ont été prévus par EURODICAUTOM[2]. Le choix de cette base se justifie d'abord par son approche notionnelle, ensuite par le fait qu'elle sert de standard à d'autres logiciels et, finalement, par le souci de notre Institut de développer un logiciel qui soit cohérent par rapport aux instruments que peuvent déjà consulter nos étudiants traducteurs.

L'extrait qui figure ci-dessous donne une idée des champs disponibles dans la fiche. La notion « vergue » ici présentée appartient à la mâture des navires. Ce microdomaine propre à la marine a été décrit à l'aide de l'actuelle version de *Termisti* parce qu'il correspond au centre d'intérêt d'un membre de l'équipe et que son caractère très structuré en fait un champ d'expérimentation particulièrement fécond.

Relations notionnelles

Nous l'avons affirmé, l'un de nos objectifs est d'approfondir la démarche notionnelle et de découvrir des principes de structuration de la connaissance. Les notions sont structurées sous la forme d'arborescences, non seulement pour enrichir la consultation, mais aussi pour permettre, dans l'avenir, une gestion du savoir assistée par ordinateur.

Établir des liens

Dans la version actuelle, on peut donc établir des relations entre les notions (non les termes) selon les principes de l'École de Vienne, comme cela a déjà le mérite de se

2. *Cf.* EURODICAUTOM, *Manuel de l'alimentateur*.

ÉCRAN 2

F2 = Sauvegarde F3 = Terme synonyme F4 = Terme équivalent F1 = Aide

Fiche terminologique

N° 0000245 Date : 25/04/91 Nom : MVC

Terme : vergue

Gram. : NF Fiabilité : 5 Dom. Lenoch : TR Dom. Eurodicautom : CNB69

Contexte

Comme les mâts et le beaupré, les vergues sont constituées par des virures de tôle (demi-coquilles rivetées), formées de deux troncs de cône assemblés par leur base; elles ont, en beaucoup plus allongé, la forme des anciennes

Note

Caractéristique

Définition

Les vergues sont des espars* généralement cylindriques, mais effilés à leurs extrémités; elles sont placées en croix sur l'avant des mâts et se désignent par le nom des voiles qui y sont enverguées; par exemple : sur la vergue de misaine est enverguée la misaine; sur la grand'vergue, la grand-voile; sur les vergues de hune, les huniers.

Rentrer la notion **Termisti — FR → EN**

faire dans le logiciel *MC4* conçu à l'université de Clermont-Ferrand. Dans la continuité de WÜSTER et conformément à la norme ISO[3], on distingue donc des relations hiérarchiques (espèce-genre et partie-tout) et non hiérarchiques, ici dites *coordonnées* (par exemple, les relations spatiales devant-derrière ou dessus-dessous, la relation cause-effet, etc.).

Attentifs aux développements de l'intelligence artificielle, on a dès à présent tenté d'exprimer ces liens par des prédications qui permettent une grande souplesse d'accès, notamment par une inversion de la perspective. Par exemple, si la notion X est située devant la notion Y, on prévoira le prédicat à deux arguments : *X est devant Y* et sa formulation inverse *Y est derrière X*.

3. ISO 704, *Principes et méthodes de la terminologie*, 1987, p. 3-4.

RELATIONS HIÉRARCHIQUES

	prédication	**prédication inverse**
Générique :	X peut être un Y	Y est un X
Partitive :	X comporte un Y	Y est une partie de X

où X est une notion superordonnée et Y, une notion subordonnée

RELATIONS COORDONNÉES

Prédication	**Prédication inverse**
X est devant Y	Y est derrière X
X est au-dessus de Y	Y est en dessous de X
X est la cause de Y	Y est dû à X
X croise Y	Y croise X
...	

où X et Y sont des notions coordonnées

Passer par une prédication nous semble, en effet, indispensable si l'on veut procher arriver à établir le réseau à partir du savoir contenu dans les définitions et les contextes. En outre, la tentative d'une description exhaustive de micro-glossaires nous a appris que les ouvrages terminographiques fournissaient très rarement des données sémantiques pour les notions hyponymes que la hiérarchie espèce-genre situe au plus bas de l'arborescence et que l'on appelle parfois « feuilles de l'arbre ». On s'est rendu compte que toute tentative de définition de ces notions aboutissait à coordonner les prédications qui situent lesdites notions dans l'arborescence. À cet égard, l'établissement de liens apparaît comme une activité éminemment définitoire.

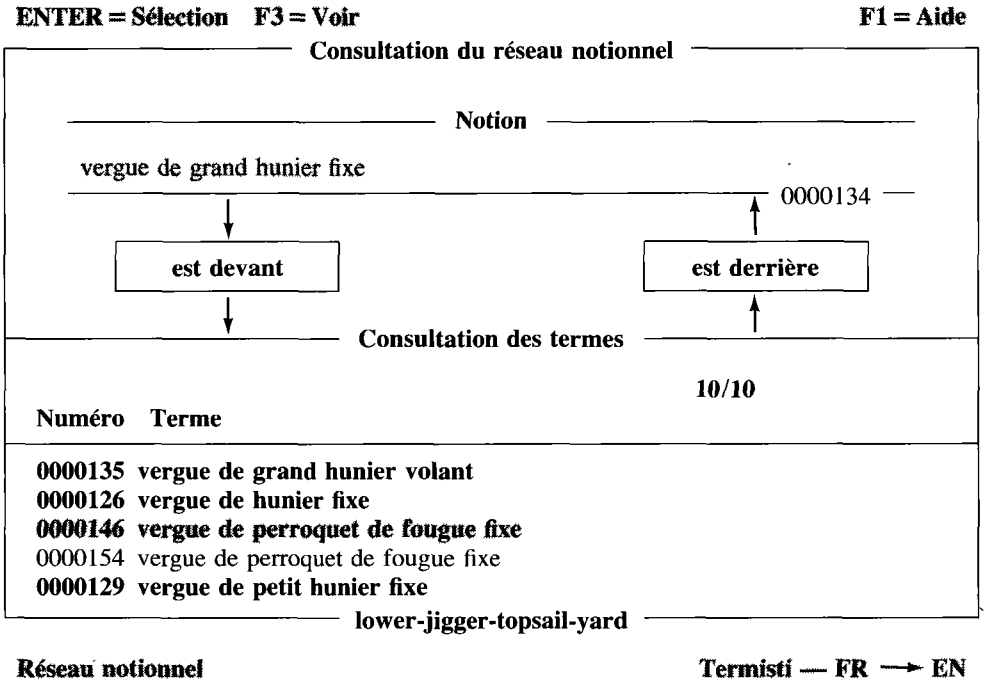
Consultation du Réseau

Ainsi, la notion *vergue de grand hunier fixe* n'est jamais définie dans les ouvrages terminographiques consacrés au vocabulaire de la marine marchande; tout au plus est-elle mentionnée sous la notion hyperonyme *vergue*. En terminologie, tenter de définir ce qu'est une vergue de grand hunier fixe, revient à préciser que cet objet est tout à la fois :

un type de — *vergue de grand hunier*
— *vergue de hunier fixe*
situé — en dessous de la *vergue de grand hunier volant*
— au-dessus de la *grand-vergue*
— derrière la *vergue de petit hunier fixe*
— devant la *vergue de perroquet de fougue fixe*

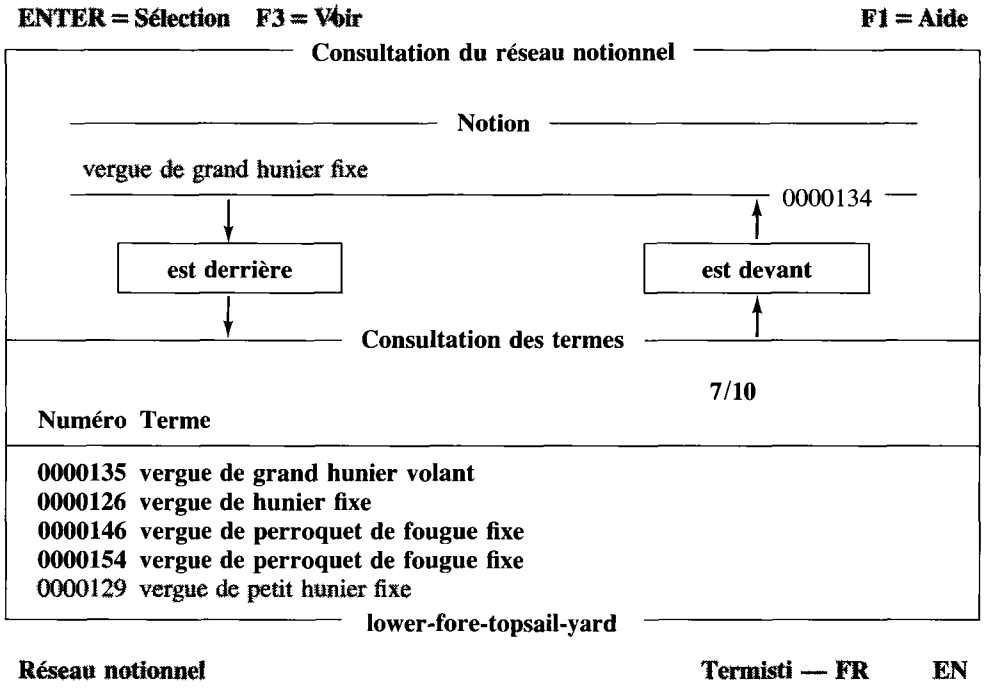
Lorsqu'on interroge *Termisti* à propos des liens qui unissent cette notion aux autres, on obtient précisément une telle réponse, illustrée par les écrans présentés ci-dessous. Bien sûr, on peut demander des informations sur chacune des notions liées et voyager dans l'arborescence; par exemple, on pourrait remonter la filiation espèce-genre pour trouver un hyperonyme défini, puis demander quelles en sont toutes les parties, etc.

ÉCRAN 5



Dans sa version expérimentale, *Termisti* apparaît donc comme un gestionnaire de notions terminologiques établies en base de données et entre lesquelles le terminologue établit des relations dans l'idée de mieux informer l'utilisateur.

ÉCRAN 6



Conclusion

La définition des termes organisés en systèmes structurés correspondant à une organisation notionnelle doit nous amener à progresser dans la compréhension de la dénomination et de la communication de savoirs spécialisés.

1. S'il est vrai qu'une terminologie scientifique ou technique est toujours insuffisante, nous croyons aboutir à une terminologie aisément mise à jour dans les sous-domaines concernés, qui peut servir les spécialistes comme les traducteurs en reflétant la pluralité mouvante des systèmes de connaissances.
2. Par l'observation des notions, de leurs relations et de leur désignation dans diverses langues, on espère approfondir la connaissance des vocabulaires de spécialité et pouvoir préciser aisément lors d'une démarche néologique, le caractère homogène ou hétérogène de structures de dénomination.
3. Nous souhaitons enfin surtout répondre à des exigences actuelles par la construction de micro-glossaires adaptés aux besoins précis des traducteurs, alliant convivialité et conceptualisation.

S'il est vrai que la terminologie se référant à l'organisation conceptuelle a des effets formateurs[4], son étude approfondie doit aboutir soit à la fin de la mythologie des arborescences telle que l'entretiennent aujourd'hui certaines pratiques terminologiques, soit à l'alimentation limitée, pertinente mais multiple du poste de travail du traducteur. Dans cette seconde hypothèse, nous pourrions parler de contribution à une « écologie traductionnelle ».

4. D. BLAMPAIN, « La formation en terminologie des traducteurs et interprètes au niveau universitaire », *Terminologies nouvelles* 2, déc. 89, p. 11-13.

27

Des fichiers terminologiques aux bases de connaissances

Bruno De Bessé

École de traduction et d'interprétation, Université de Genève, Genève, Suisse

Introduction

Pour obtenir une traduction de qualité, il convient de disposer d'outils terminologiques nombreux, variés, exhaustifs, adéquats et fiables. On estime en effet que la recherche terminologique peut représenter plus de 50 pour cent du temps consacré à la traduction dans les domaines scientifiques et techniques, juridiques et économiques.

Le recours à l'informatique s'est donc imposé comme la seule solution capable de permettre la rapidité d'entrée des informations terminologiques, leur regroupement selon les besoins des utilisateurs, ainsi que leur tenue à jour constante et rapide. Ainsi, depuis quelques années, la profession de traducteur connaît-elle une évolution importante. Le traducteur fait de plus en plus appel à l'informatique pour le seconder. Il abandonne progressivement son crayon et sa gomme, son dictaphone, sa machine à écrire, ses dictionnaires traditionnels, le fichier qu'il a constitué dans une boîte à chaussures. Aujourd'hui le traducteur rédige à l'aide du traitement de texte, consulte les banques de terminologie et les dictionnaires automatisés, et crée ses propres outils terminologiques. On assiste à la création d'un véritable poste de travail du traducteur, combinant ces différentes aides informatiques.

Les produits

Les grandes banques

Les grandes banques de terminologie ont été les premières aides informatiques utilisées par les traducteurs.

Les grandes banques font leur apparition entre 1963 et 1975 : DICAUTOM en 1963 (banque de la Commission des Communautés européennes) qui devient EURODICAUTOM en 1973, LEXIS en 1966 (banque de l'Office fédéral des langues (*Bundessprachenamt*) du ministère de la Défense du gouvernement de la République fédérale d'Allemagne), TEAM en 1967 (banque de la société SIEMENS), BTUM en 1970 (banque de l'Université de Montréal) qui sera reprise en 1975 par le Bureau des traductions du gouvernement du Canada et deviendra TERMIUM, BTQ en 1974 (banque de terminologie du gouvernement du Québec).

On s'est très vite rendu compte que seule l'informatique offrait la possibilité de gérer des fonds terminologiques particulièrement importants qui ne sont plus maîtrisables par les moyens traditionnels.

Une banque de terminologie est un dictionnaire informatisé. Le dictionnaire traditionnel est un recueil d'articles; la banque de terminologie est une collection de fiches. La fiche terminologique d'une banque est tout à fait comparable à l'article d'un dictionnaire. Elle contient habituellement les informations suivantes : une entrée, une définition, l'indication de domaine, des synonymes s'il y a lieu, les équivalents de l'entrée dans une ou plusieurs langues, et, le cas échéant, un exemple d'emploi (le contexte), des marques d'usage, des informations encyclopédiques.

Plusieurs grandes banques de terminologie fonctionnent actuellement : la banque de terminologie du Canada (TERMIUM), la banque de terminologie du Québec (BTQ), la banque de terminologie de la Commission des Communautés européennes (EURODICAUTOM), la banque de terminologie de l'Office fédéral des langues, dépendant du ministère de la Défense de la République fédérale d'Allemagne (LEXIS), la banque de terminologie de la société SIEMENS (TEAM).

Certaines banques sont disponibles sur CD-ROM.

Brève présentation de deux banques de terminologie

*EURODICAUTOM

EURODICAUTOM est un système de gestion de terminologie multilingue créé par la Commission des Communautés européennes, pour aider les traducteurs et les fonctionnaires des institutions européennes dans la recherche d'équivalences terminologiques dans les langues officielles de la Communauté européenne.

Les langues traitées dans EURODICAUTOM sont l'allemand, l'anglais, le danois, l'espagnol, le français, le grec, l'italien, le néerlandais et le portugais.

EURODICAUTOM contient 470 000 fiches terminologiques (chaque fiche contenant en général cinq à six langues) et 120 000 abréviations.

Les principaux domaines sont : administration publique et privée, automatisation, Communautés européennes, commerce, économie, électrotechnique, environnement, finances, industries chimiques, droit, sciences médicales et biologiques, mécanique générale, industrie minière, arts graphiques, physique, sidérurgie et fonderie, techniques et industries, transports, travail.

EURODICAUTOM est accessible à toute personne qui en fait la demande. L'accès se fait par les réseaux télématiques et le serveur ECHO. L'accès est gratuit. Les utilisateurs ne paient que les frais de télécommunications.

Seule l'interrogation est possible. L'utilisateur n'a pas la possibilité de saisir lui-même de nouveaux termes.

EURODICAUTOM est incompatible avec les traitements de texte. Il n'est donc pas possible d'insérer les termes dans le traitement de texte.

*TERMIUM

TERMIUM est la banque de terminologie du gouvernement du Canada. Elle est essentiellement utilisée par ses fonctionnaires aux niveaux fédéral et provincial.

Les langues traitées sont l'anglais et le français. Quelques collections terminologiques sont disponibles en espagnol et allemand.

TERMIUM contient environ 900 000 fiches.

Les domaines principaux sont : sciences, industries, transport et manutention, administration et gestion, économie, sciences humaines, médecine, écologie, défense, postes et télécommunications, électricité, agriculture.

TERMIUM n'est pas accessible en ligne au grand public en dehors du Canada, à l'exception des organisations internationales et des administrations nationales. Il est toutefois possible de s'abonner à la version CD-ROM de cette banque.

Seule l'interrogation est possible. L'utilisateur n'a pas la possibilité de saisir lui-même des fiches nouvelles. La banque offre la possibilité de signaler les « termes absents » (à inclure dans la banque).

TERMIUM est incompatible avec les traitements de texte. Il n'est donc pas possible d'insérer des termes dans le traitement de texte.

Les banques moyennes

Les banques moyennes se caractérisent par des moyens informatiques modestes, un nombre limité de termes, une spécialisation dans un domaine particulier et un nombre d'utilisateurs restreint. De nombreuses banques moyennes ont été construites en Europe, en particulier par les grandes entreprises.

Les CD-ROM

Plusieurs recueils terminologiques et lexicographiques sont disponibles sur CD-ROM.

On trouvera ci-après la présentation de quelques-uns de ces outils.

TERMIUM

Description :

Banque de terminologie bilingue (anglais-français) couvrant tous les domaines.

Contenu :

900 000 fiches (anglais-français) extraites de la banque de terminologie du Secrétariat d'État du Canada.

Fonctionnement :

- Pas résident.
- Interface utilisateur en français ou en anglais.
- Index des termes anglais, des termes français, des mots anglais et des mots français.
- Possibilité d'interroger à partir des mots contenus dans la définition, dans le contexte ou dans la note.
- Possibilité de faire défiler l'index.
- Possibilité d'utiliser les opérateurs booléens pour combiner plusieurs clés de recherche.
- Possibilité de copier des fiches sur le disque dur, sur une disquette ou de les imprimer.
- Possibilité de créer différents formats d'affichage.
- Possibilité d'utilisation en réseau à certaines conditions.
- Mise à jour : trois fois par an.

TERMDOK 89

Description :

Banque de terminologie technique multilingue (langues : suédois, anglais, français, allemand, espagnol, russe, danois, norvégien, finnois).

Contenu :

Glossaires du Centre suédois de terminologie technique; normes suédoises; extraits de documents préparés par le Centre suédois de terminologie technique; banque de terminologie du Centre norvégien de terminologie technique; banque finlandaise de normes et glossaires; 100 000 fiches (anglais-français) extraites de TERMIUM; 47 000 fiches (anglais-français) extraites de NORMATERM.

Fonctionnement :

- Possibilité d'installation comme programme résident.
- Trois langues de travail (anglais, français, suédois).
- Logiciels d'interrogation et de formation sur le CD-ROM.
- Possibilité d'utiliser le résultat d'une recherche pour la recherche suivante.
- Possibilité de limiter la recherche à quelques bases seulement.
- Possibilité de combiner plusieurs clés de recherche.
- Possibilité de faire défiler l'index.
- Possibilité d'afficher jusqu'à quatre fiches à la fois.

- Affichage par défaut de tous les mots commençant par les caractères du mot clé.
- Possibilité, grâce à la fonction calepin, de sauvegarder temporairement jusqu'à 2000 caractères, qui peuvent ensuite être imprimés, sauvegardés sur disquette ou transférés dans un programme de traitement de texte.
- Pas de possibilité d'utilisation en réseau.
- Mise à jour : annuelle.

Le Robert électronique

Description :

Ensemble d'outils d'aide à la rédaction fondé sur la base de données servant à préparer la nouvelle édition du *Grand Robert de la langue française*.

Contenu :

Nomenclature de 100 000 entrées; dictionnaire de 80 000 articles; anthologie littéraire de 160 000 citations; dictionnaire analogique comprenant un million de synonymes, homonymes, dérivés, composés et antonymes; liste d'abréviations et de signes conventionnels; annexe bibliographique; conjugaison complète de tous les verbes.

Fonctionnement :

- Possibilité d'installation comme programme résident.
- Compatible avec les principaux logiciels de traitement de texte.
- Possibilité d'effectuer soit une recherche abrégée (structure et sens principaux), soit une recherche détaillée (pour chaque sens : définition principale, exemples, éléments de phraséologie, renseignements analogiques, définitions secondaires, remarques et commentaires grammaticaux).
- Possibilité d'interroger par chaîne de caractères.
- Possibilité d'insérer les éléments sélectionnés dans le fichier de travail.
- Possibilité d'interroger à partir d'une graphie fautive.
- 450 000 formes verbales.
- Vérificateur orthographique.

ZYZOMYS

Description :

Dictionnaire électronique du français sur disque optique compact.

Contenu :

Dictionnaire de notre temps (Hachette); *Dictionnaire des synonymes* (Hachette); *Atlas pratique* (Hachette). 400 000 formes de la langue actuelle à partir de 72 000 entrées dont 21 000 noms propres; plusieurs dizaines de cartes en couleurs.

Fonctionnement :

- Pas résident (à l'exception du vérificateur d'orthographe).
- Possibilité cependant de récupérer l'information dans le traitement de texte.
- Possibilité d'interroger par mot vedette.
- Possibilité d'interroger à partir des mots ou des chiffres contenus dans la définition.
- Possibilité d'interroger par chaîne de caractères.
- Possibilité d'interroger par date.
- Possibilité de visualiser les tableaux de conjugaison.
- Possibilité d'afficher les cartes géographiques.
- Possibilité d'utilisation en réseau.
- Mise à jour : annuelle.

The Original Oxford English Dictionary

Description :

Dictionnaire unilingue.

Contenu :

Oxford English Dictionary en 12 volumes; pour chaque entrée : étymologie, citations et renseignements divers; 42 millions de mots; 252 259 entrées; 1 860 000 citations.

Fonctionnement :

- Pas résident.
- Possibilité d'interroger par mot clé, par étymologie ou par domaine.
- Possibilité de retrouver des citations par la date, par le nom de l'auteur, par des mots y figurant ou par le texte de la citation.
- Possibilité de sauvegarder des formats de recherche.
- Possibilité d'utiliser des caractères spéciaux, en particulier pour l'étymologie.
- Possibilité d'utilisation en réseau.

Merriam-Webster's Ninth New Collegiate Dictionary

Description :

Contenu du *Merriam-Webster's Ninth New Collegiate Dictionary* (articles, illustrations et prononciation enregistrée de toutes les entrées).

Contenu :

160 000 articles.

Fonctionnement :

- Pas résident.
- Enregistrement numérique de la prononciation américaine standard.
- Possibilité d'affichage en caractères normaux ou larges.

- Illustrations du dictionnaire inclues dans le CD-ROM.
- Noms propres compris dans la nomenclature.

The New Grolier Electronic Encyclopedia

Description :

Encyclopédie unilingue.

Contenu :

Academic American Encyclopedia en 21 volumes (33 000 entrées; 9 millions de mots).

Fonctionnement :

- Pas résident.
- Possibilité d'ouvrir jusqu'à 10 fenêtres contenant des informations.
- Possibilité d'étendre la recherche par des liens automatiques.
- Possibilité de combiner plusieurs clés de recherche.
- Possibilité d'afficher l'historique de la recherche.
- Possibilité de tronquer les clés de recherche.
- Possibilité de transférer les informations sur une disquette ou de les imprimer.
- Possibilité d'utilisation en réseau (version MS-DOS).

CD-WORD 12 + 1

Description :

Ensemble de 18 dictionnaires bilingues ou multilingues en 12 langues.

Contenu :

Harrap's Shorter French-English, English-French Dictionary; *Harrap's French-English, English-French Dictionary of Data Processing*; *Harrap's French-English, English-French Science Dictionary*; *Harrap's Concise German-English, English-German Dictionary*; *Harrap's Business Dictionary : English-French, French-English*; *Compact Dictionary of Exact Science and Technology, English-German, German-English*; etc.

Fonctionnement :

- Possibilité d'installation comme programme résident.
- Compatible avec les principaux logiciels de traitement de texte.
- Possibilité d'interroger des couples de langues sans l'anglais, en utilisant l'anglais comme langue pivot.
- Possibilité de choisir les dictionnaires et les langues.
- Possibilité d'interroger à partir de vedettes, d'entrées secondaires ou d'exemples.
- Possibilité d'interroger en inversant la langue source et la langue cible, et d'interroger sur les constituants d'un mot composé.
- Possibilité d'effectuer une recherche sur une liste de mots.
- Pas de possibilité d'utilisation en réseau.

Les logiciels

Il existe un grand nombre de logiciels qui permettent aux traducteurs de construire leur propre banque.

Voici une courte présentation de quelques logiciels choisis parmi les plus connus.

AQUILA

Aquila est un logiciel de gestion et d'exploitation de base de données terminologiques multilingues qui peut être utilisé parallèlement à un traitement de texte, mais qui n'est pas résident. Il gère deux langues simultanément et peut accueillir quatorze langues en implantation. L'organisation et la recherche des données se fait de manière pyramidale. Les liens entre les différentes langues se font automatiquement.

Il s'agit d'un programme monoposte avec une interface utilisateur en français, en allemand, en anglais, en italien, en espagnol.

La fiche a 13 champs structurés et peut contenir 642 caractères : terme dans la langue source (78 caractères), nature (3 caractères), genre (un caractère), source (50 caractères), terme dans la langue cible (78 caractères), nature (3 caractères), genre (un caractère), source (50 caractères), domaine (3 caractères), secteur (3 caractères), maison (3 caractères), définition (135 caractères), note technique (234 caractères).

Plusieurs glossaires sont disponibles, notamment dans les domaines suivants : électronique, télécommunications, informatique, aéronautique, droit-économie, technologies, mécanique.

Le logiciel est compatible avec tous les traitements de texte connus.

PROFILEX

Profilex est un logiciel pour dictionnaire électronique qui permet entre autres de rechercher des termes et de les transférer dans un texte, de créer ou de modifier une fiche sans interrompre le travail dans le traitement de texte, d'effectuer des restrictions de domaine.

Le logiciel est conçu pour la confection de dictionnaires unilingues, bilingues ou multilingues.

Il existe en version monoposte et en version multiposte avec une interface utilisateur en allemand, en français et en anglais.

Les champs sont de longueur variable. Quatre cent trente caractères sont disponibles pour chaque langue. Il est possible d'annexer au champ principal un ou plusieurs champs complémentaires. Les synonymes sont séparés par des points-virgules. Les autres informations figurent entre parenthèses.

Le nombre d'entrées dans le dictionnaire n'est limité que par la capacité de mémoire du disque dur.

Tout fichier terminologique sur support magnétique peut être mis en mémoire à l'aide de programmes de conversion.

Le logiciel est compatible avec tous les traitements de texte.

SUPERLEX

Superlex comprend trois modules : *Superlex-Glossarium*, *Superlex-Windows* et un programme d'impression.

Superlex-Glossarium est un éditeur pleine page qui permet de créer et de modifier des dictionnaires.

Superlex-Windows est résident et peut être appelé à tout moment pour afficher un terme et l'insérer dans une traduction. Il est possible de créer de nouvelles fiches ou d'ajouter de nouveaux termes sans quitter le traitement de texte. Langue source et langue cible peuvent être inversées par une touche sans qu'il y ait double sauvegarde.

Le programme d'impression comprend un pilote d'impression permettant d'obtenir des sorties en deux langues.

Le logiciel existe en version monoposte et en version multiposte.

La fiche est bilingue. Chaque langue dispose d'un champ. La fiche ne peut contenir plus de 304 caractères. Un champ est constitué de 4 lignes de 38 caractères : la première ligne contient le terme; les autres lignes contiennent les informations sur le terme. Des champs bilingues peuvent être annexés si cela est nécessaire.

Le logiciel peut gérer jusqu'à 90 dictionnaires contenant chacun 65 000 fiches bilingues.

La conversion de données est possible par l'interface ASCII.

Le programme est compatible avec la plupart des traitements de texte.

TERMEX

Termex est un logiciel de gestion de fichiers terminologiques qui offre la possibilité d'ajouter, de modifier ou de supprimer des entrées, à partir d'un traitement de texte. Des programmes utilitaires permettent de fusionner plusieurs dictionnaires et d'en extraire des sous-ensembles.

Le logiciel existe en version monoposte et en version multiposte. Il fonctionne avec les caractères romains et les caractères cyrilliques, et a une interface utilisateur en allemand, français, anglais.

Il n'y a pas de limitation du nombre de champs. La longueur d'une fiche est variable et peut aller jusqu'à 50 lignes.

Le nombre d'entrées maximal est d'un million de fiches, les seules limites sont celles du système d'exploitation, du disque dur et des disquettes.

Des programmes utilitaires permettent de convertir des collections terminologiques vers ou à partir de fichiers standard de traitement de texte.

Des dictionnaires sont disponibles dans plusieurs domaines : aéronautique, économie, informatique, commerce international, etc.

Le logiciel est compatible avec la plupart des traitements de texte.

TERM-PC

Term-PC est un logiciel pour le travail terminologique professionnel qui permet de créer, de gérer et d'exploiter des fichiers terminologiques unilingues ou multilingues, et de récupérer des fichiers terminologiques et de les réviser.

Le logiciel existe en version monoposte (MS-DOS) et en version multiposte (SINIX) avec une interface utilisateur en allemand, anglais ou espagnol.

Le format standard du masque offre 10 champs par langue : terme, abréviation, définition, contexte, source, mot clé, domaines et deux champs libres. La fiche peut contenir 4000 caractères.

Le format peut être modifié, et en particulier agrandi, selon les besoins de l'utilisateur.

Le nombre d'entrées n'est limité que par la capacité du disque dur.

La conversion de fichiers au format ASCII, ou l'inverse, par disquette ou téléchargement est possible.

Term-PC est compatible avec TEAM, banque de terminologie de la société Siemens fonctionnant sur gros système.

Des dictionnaires sont disponibles dans plusieurs domaines : informatique, mécanique, technique des câbles, etc.

Ce système n'offre pas la possibilité d'insertion dans le traitement de texte. Un logiciel supplémentaire est nécessaire pour la connexion entre *Term-PC* et les logiciels de traitement de texte.

TERMTRACER

TermTracer est un logiciel qui donne accès à des dictionnaires bilingues pendant l'utilisation du traitement de texte. Il est possible à tout moment de consulter les différents dictionnaires bilingues, de les modifier ou même d'en créer de nouveaux. Le programme permet l'insertion des équivalents dans le traitement de texte. Des utilitaires permettent d'inverser les langues dans un dictionnaire, de fusionner et de comparer deux dictionnaires.

Le logiciel existe en version monoposte et en version multiposte.

La fiche se compose de cinq champs dont les dimensions sont fixes, et comprend 261 caractères : terme (maximum 64 caractères), équivalent (maximum 64 caractères), forme (maximum 5 caractères), usage (maximum 64 caractères), autres informations (maximum 64 caractères).

La taille des dictionnaires est uniquement limitée par la mémoire disponible.

Il est possible de convertir des glossaires au format ASCII, ou l'inverse. Des dictionnaires sont disponibles dans plusieurs domaines : informatique, économie et commerce, médecine, droit.

Le programme est compatible avec la plupart des traitements de texte.

Et les autres...

En outre, de nombreux utilisateurs ont mis au point des logiciels pour la gestion de leurs données terminologiques. Ces produits sont destinés à l'usage personnel de leur auteur et ne sont pas en général commercialisés.

Évaluation

Les grandes banques

Pour le traducteur, une banque de terminologie a théoriquement l'avantage d'être exhaustive, d'être à jour, d'être accessible en tout temps et en tout lieu.

L'utilisation des banques permet-elle au traducteur d'augmenter à la fois la quantité et la qualité de son travail? Difficile à quantifier avec précision, le gain de temps ainsi réalisé par le traducteur se situerait dans le meilleur des cas entre 10 et 20 pour cent.

Les critiques formulées par les traducteurs qui consultent ces banques sont bien connues. Ils leur reprochent le trop grand nombre de réponses fournies. Les utilisateurs souhaiteraient avoir la possibilité d'affiner les questions et de limiter le nombre de réponses non pertinentes. Ils regrettent le nombre important de questions restées sans réponses. Les termes recherchés ne se trouvent pas dans la banques. Les taux de réponse sont extrêmement variables d'un domaine à l'autre, allant de 50 pour cent à 75 pour cent. Ils considèrent que leurs contenus ne sont pas assez spécialisés. Ils trouvent que leur mise à jour est lente.

L'utilisateur n'a pas de possibilité d'alimenter lui-même ces banques.

À l'intérieur de l'organisme qui dispose d'une banque, la banque de terminologie peut être interrogée en mode conversationnel ou en différé. Pour les utilisateurs extérieurs, seule l'interrogation en mode conversationnel est possible.

Pour interroger une banque de terminologie l'utilisateur extérieur doit passer par les réseaux télématiques. La procédure d'accès aux réseaux est en général considérée comme laborieuse. Les coûts d'utilisation dépendent de la durée de la communication et du volume des données transportées. Leur interrogation est en général considérée comme coûteuse.

Pendant l'accès au réseau, le traducteur ne peut utiliser son traitement de texte.

Les banques moyennes

Les CD-ROM

Le CD-ROM permet de stocker une très grande quantité de données. Un disque peut contenir 250 000 pages dactylographiées. *Le Grand Robert* en 10 volumes tient sur un seul disque.

Grâce aux CD-ROM, le traducteur peut non seulement consulter les banques de terminologie ou les dictionnaires, à faible coût, mais encore intégrer directement le résultat de ses recherches dans la traduction qu'il est en train de préparer.

En outre, les banques de terminologie sur CD-ROM offrent des possibilités d'interrogation qui ne sont pas disponibles en ligne.

Les CD-ROM présentent également l'avantage de pouvoir être interrogés à tout moment, sans avoir à se soucier des encombrements téléphoniques ou des problèmes de réseaux, etc.

L'interrogation du CD-ROM s'effectue à l'aide d'un logiciel qui est livré avec chaque produit. Il n'y a malheureusement pas encore de langage commun d'interrogation, ce qui signifie qu'il faut apprendre un langage d'interrogation pour chaque CD-ROM.

Le CD-ROM est un support qui permet seulement la lecture des informations qu'il contient, mais pas la saisie de nouvelles données.

Les CD-ROM ne sont pas toujours régulièrement mis à jour.

À signaler enfin qu'avec la multiplication du nombre des CD-ROM, la distinction faite traditionnellement entre banques de terminologie et CD-ROM prend une dimension nouvelle. On retrouve en effet, au-delà du support, l'opposition traditionnelle entre dictionnaires de mots et dictionnaires de choses.

Les logiciels

Tout traducteur dispose de son propre fichier terminologique, qu'il garde précieusement et un peu jalousement. Ce fichier contient des termes rares, des tournures et expressions particulières, et surtout des termes nouveaux qui ne figurent encore dans aucun dictionnaire, et qui ont été trouvés au prix d'une longue et laborieuse recherche.

L'informatique peut simplifier la gestion de ces ressources. Mais il ne saurait être question de saisir ces glossaires à l'aide d'un traitement de texte. Un système de gestion de base de données (SGBD) serait sans doute plus approprié pour ce travail, s'il n'existait pas déjà sur le marché de nombreux programmes spécialement conçus pour la gestion de la terminologie.

Les outils informatiques ne manquent pas dans ce domaine et l'utilisateur a plutôt l'embarras du choix. Depuis quelque temps en effet, on assiste à une multiplication des produits, qui peuvent être évalués selon différents critères.

Ces programmes sont-ils résidents ou non résidents, monopostes ou multipostes? Ces logiciels ont-ils été conçus comme des logiciels de banque de données? Comment les champs sont-ils organisés? Il convient également de prendre en compte le mode de gestion de la mémoire et les possibilités d'échange de données.

Le logiciel est-il utilisable en même temps que le traitement de texte, ou bien l'utilisateur doit-il aller d'un programme à l'autre? Dans ce cas, il n'est pas possible de transférer automatiquement un terme dans une traduction, ni d'alimenter le fichier terminologique pendant l'utilisation du traitement de texte.

Toutefois, les programmes à fonctionnement multitâche, de même que les nouveaux systèmes d'exploitation, permettent de pallier en partie cet inconvénient grâce au système des fenêtres. Il est ainsi possible de travailler avec un traitement de texte dans une fenêtre et avec un logiciel de terminologie dans l'autre.

La plupart des logiciels de terminologie sont conçus pour fonctionner en version monoposte ou en version multiposte. Ils répondent ainsi aux besoins du traducteur indépendant comme à ceux des entreprises.

Quelques producteurs de logiciels ont sans aucun doute trouvé leur inspiration dans les logiciels de banques de données. Mais tous les logiciels ne sont pas conçus comme des logiciels de banque de données.

La conception du programme a des conséquences importantes sur l'interrogation. Peut-on interroger les fichiers à partir d'un domaine, d'une source, d'un nom d'auteur, etc.? Le terme est-il la seule clé de recherche?

La gestion des synonymes varie d'un système à l'autre. Le plus souvent, l'utilisateur doit saisir les synonymes sur des fiches distinctes.

Rares sont les programmes qui permettent la gestion de fiches comportant trois langues ou plus. Certes, il est possible d'implanter plusieurs langues, mais l'utilisateur ne pourra interroger qu'en mode bilingue. Parfois, le lien entre deux langues se fait par l'intermédiaire d'une langue pivot.

L'utilisateur doit-il remplir obligatoirement tous les champs? Peut-il en fixer le nombre et la longueur?

Dans certains logiciels l'espace est fortement structuré. Les inconvénients sont évidents. Certains champs sont inutiles. En revanche d'autres sont insuffisants. L'organisation de la fiche peut être relativement complexe et compter dix champs ou plus, mais elle peut aussi être très rudimentaire. La longueur maximale des champs varie sensiblement d'un système à l'autre.

Certains produits semblent avoir été conçus indépendamment des besoins des utilisateurs. Ainsi, dans certains programmes les champs sont-ils prédéterminés au départ et ne laissent pas de liberté au traducteur. Dans ce cas l'utilisateur n'a la possibilité ni d'ajouter, ni de supprimer un champ, ni de modifier sa dimension. Parfois, le nombre de caractères offert dans certains champs est tout à fait insuffisant. En outre, la méthode qui consiste à regrouper dans un seul champ des informations aussi différentes que les données grammaticales, les abréviations, le domaine, la définition, les notes, le contexte n'est pas très satisfaisante.

À l'opposé, d'autres programmes offrent une liberté de structuration totale. L'utilisateur se trouve alors en face d'un écran vide. Il fixe lui-même le nombre et la longueur des champs et crée un masque de saisie.

Ces produits se caractérisent par leur absence de structuration. Ils laissent au traducteur la liberté totale d'organiser et de structurer ses champs. Ce manque de structuration rend difficile la coordination en cas d'utilisation par plusieurs personnes. En général, l'absence de champs complique la gestion des informations terminologiques et inquiète l'utilisateur.

D'autres proposent une semi-liberté de structuration. Dans ce cas tous les renseignements complémentaires sont ajoutés assez librement après la vedette, soit entre parenthèses, soit sur des lignes supplémentaires.

Certes, la liberté de structuration permet de prendre en compte les besoins de l'utilisateur. Mais elle implique néanmoins une très grande discipline de leur part, s'ils veulent avoir un produit cohérent, surtout s'ils sont plusieurs à travailler avec le même outil.

Bref, le traducteur ne veut ni d'un système trop rigide ni d'un système trop invertébré.

La gestion de la mémoire est un point important, qui est lié à la conception des champs. Les champs sont-ils sauvegardés intégralement, qu'ils soient remplis ou non? Le programme comprime-t-il l'espace qui n'a pas été utilisé? Ces différences de conception ont des conséquences sur le matériel.

On constate qu'il n'existe pas de logiciel pour banque de terminologie de taille moyenne, compatible avec la majorité des traitements de texte, et répondant vraiment aux besoins terminologiques de la grande majorité des traducteurs, qui souhaitent disposer d'un système se caractérisant par une rigueur de conception et de construction, une cohérence d'organisation, une souplesse de gestion, une facilité d'utilisation.

Ces considérations n'ont, semble-t-il, pas échappé aux éventuels acheteurs qui se font plutôt rares. Il serait intéressant de connaître le nombre d'exemplaires vendus pour se faire une idée exacte du succès remporté par chaque produit et de l'intérêt qu'ils ont suscité. Les producteurs de logiciels se gardent bien de fournir la moindre indication à ce sujet. Mais il semble bien que les utilisateurs, qui ont toujours le dernier mot, ne se soient pas laissé abuser.

La solution consiste peut-être, comme le proposent certains, à former les traducteurs à l'informatique, pour qu'ils soient en mesure de mettre au point leurs propres outils. Après la profession de linguiste-informaticien, on peut imaginer celle de traducteur-informaticien. Il n'est pas sûr que la plupart des traducteurs soient disposés à jouer ce rôle. La traduction est un métier, l'informatique un autre. Même si les complexes à l'égard de l'informatique ne sont pas de mise, les cours de programmation n'ont pas leur place dans les écoles de traduction.

Conception

La fiche constitue l'unité de base de toutes les banques de terminologie.

La conception des dictionnaires et des banques de terminologie est fondamentalement la même. Ces produits sont constitués d'une collection d'articles ou de fiches, organisés autour d'une entrée ou d'une vedette, et qui contiennent des informations sur les choses, sur les concepts, sur les termes ou sur les mots. L'accès à ces informations se fait par l'intermédiaire des mots ou des termes figurant en entrée ou en vedette. Dans tous les cas, le passage par un mot clé est obligatoire.

Les développements récents de l'informatique permettent désormais d'interroger l'article ou la fiche par tous les mots qu'il contient. Cette possibilité constitue un progrès important et multiplie les accès aux informations disponibles. Mais la conception de l'accès reste la même : elle se fait par des mots ou par des termes.

L'informatique facilite la gestion des données lexicographiques et terminologiques. Mais elle n'a pas encore modifié la conception traditionnelle de l'utilisation des outils dans ce domaine. Elle ne fait que reproduire les procédures d'interrogation des dictionnaires en vigueur depuis leur naissance.

Aucun système fondé sur l'accès à l'information par l'intermédiaire des concepts et de leurs constituants (les caractères) n'est actuellement disponible.

Les besoins

Tous les traducteurs doivent pouvoir utiliser les dictionnaires informatisés disponibles dans le commerce et créer leurs propres dictionnaires sur ordinateur.

Le fichier terminologique du traducteur indépendant doit en outre être compatible avec son traitement de texte, se prêter au transfert de données dans le traitement de texte, pouvoir être alimenté à partir du traitement de texte et offrir des champs susceptibles d'être modifiés selon les besoins.

Le traducteur d'entreprise a les mêmes besoins que le traducteur indépendant. Mais la capacité de sa banque de terminologie doit être plus importante; elle doit en outre être interrogeable par plusieurs utilisateurs en même temps; elle doit enfin être compatible avec le réseau interne de l'entreprise et accessible en ligne à tout moment.

De plus en plus, on assiste à l'apparition de programmes permettant de constituer un véritable poste de travail du traducteur. Dans ce cas, il s'agit pour l'ordinateur d'assister le traducteur et non la traduction, au-delà de la simple utilisation en parallèle d'un traitement de texte et d'un logiciel de terminologie. Le poste de travail intégré pourrait à long terme constituer la solution d'avenir pour le traducteur.

Perspectives

Pour alimenter la réflexion sur la conception des dictionnaires de l'an 2001, une équipe de l'Université de Genève[1] a entrepris, dans le cadre d'un projet de recherche mul-

1. Bruno DE BESSÉ (chargé de cours), Michel BONJOUR (assistant), Gilles FALQUET (maître-assistant), Michel LÉONARD (professeur ordinaire), Donatella PULITANO (assistante), Jeanne SINDAYAMAZE (assistante).

tidisciplinaire associant l'École de traduction et d'interprétation, la Faculté des sciences économiques et sociales et le groupe « Bases de données » du Centre universitaire informatique, et financé par le Fonds national de la recherche scientifique, la construction d'un dictionnaire encyclopédique multilingue de concepts.

L'objectif de la recherche est de mettre au point un dictionnaire de concepts dans un domaine rigoureusement délimité. Ce dictionnaire se présentera sous la forme d'une base de données contenant tous les concepts appartenant à ce domaine, ainsi que tous les caractères qui les composent. La méthode retenue pour la description des concepts est celle qui est utilisée en terminologie. La démarche est onomasiologique. Elle procède à partir des choses, des réalités, pour aller aux concepts, puis aux mots ou aux termes.

On entend par concept toute unité de pensée constituée d'un ensemble de caractères attribués à un objet ou à une classe d'objets.

On entend par caractère chacune des qualités, des propriétés ou des relations qui composent le concept.

Le système à développer constituera une base de connaissances. Sa construction passe par une modélisation des connaissances effectuée à partir d'une analyse des concepts et de leurs caractères. Il s'agit, pour ce faire, de procéder à une véritable dissection des concepts, de décrire leur contenu, leur organisation et leurs rapports, de montrer leurs relations, de représenter leur structuration, d'établir les liens, hiérarchiques ou non, qui les unissent.

L'interrogation de ce dictionnaire se fera à partir de concepts constitués, ou à partir d'ensembles de caractères constituant des concepts ou susceptibles d'en constituer. Les associations de caractères permettront également de générer des concepts dont l'existence est virtuelle.

La base contiendra les formes linguistiques attachées aux concepts et à leurs caractères, qui joueront accessoirement le rôle de clé d'accès au dictionnaire. Les équivalents linguistiques des concepts seront donnés dans plusieurs langues. Ils seront accompagnés, le cas échéant, d'indications sur les différences de contenu des concepts selon les langues.

Traditionnellement, quand on consulte un dictionnaire ou une banque de terminologie, on interroge un mot ou un terme pour connaître sa définition. Dans le prototype qui sera développé, la démarche consistera à chercher le mot ou le terme à partir du contenu et des éléments de la définition.

Le système se proposera de répondre aux questions suivantes :

- Q. Comment s'appelle un siège à dossier, sans bras?
- R. Chaise.
- Q. Comment s'appelle en anglais un siège à dossier, sans bras?
- R. Chair.

La base de concepts constituera une banque de savoir et servira à l'acquisition de nouvelles connaissances. À l'aide des informations fournies, et notamment d'une description d'un objet ou de l'énumération de ses caractéristiques, le prototype sera en

mesure de répondre aux questions que se pose tout apprenant : Comment s'appelle l'objet qui... ? Comment cela s'appelle? Qu'est-ce que c'est?

Un tel dictionnaire provoquera une remise en cause des méthodes lexicographiques et terminologiques. Il mettra en évidence les faiblesses de ces outils et contribuera à leur amélioration.

L'informatique, qui jusqu'alors s'est contentée de faciliter la gestion de ce domaine, pourra enfin apporter une véritable contribution portant sur la conception des outils.

Plusieurs programmes seront créés pour gérer la structure, les concepts et les caractères du domaine, associer les formes linguistiques aux concepts, relier le système de concepts aux formes linguistiques de chaque langue, intégrer les différents médias, permettre toutes les formes d'interrogation du dictionnaire, utiliser la base à des fins d'apprentissage.

Le dictionnaire doit passer à l'ère informatique. L'informatique ne doit plus seulement servir d'aide à l'édition. Cette évolution doit s'accompagner d'une remise en cause et d'un changement des méthodes de production. C'est la conception même des dictionnaires qu'il convient de modifier.

L'objet principal de la recherche projetée sera de fournir un moyen de trouver les mots ou les termes, quand on a seulement l'idée des choses.

Le dictionnaire deviendra ainsi une base de connaissances, associant différents médias. Il abandonnera son rôle de mémoire passive pour devenir un outil interactif intelligent.

Conclusion

Depuis quelques années le traducteur a la possibilité d'utiliser plusieurs banques de terminologie. De nombreux logiciels de terminologie sont également à sa disposition.

Ces outils sont encore loin de lui donner toutes satisfactions. Il est nécessaire de leur apporter de nombreuses améliorations, aussi bien sur le plan informatique que sur le plan terminologique.

La conception de ces banques et de ces logiciels est fondamentalement la même. L'accès se fait par l'intermédiaire de vedettes. Les développements récents de l'informatique permettent d'interroger la fiche par d'autres champs. Cette possibilité constitue un progrès important et multiplie les accès aux informations disponibles.

Aucun système fondé sur l'accès à l'information par l'intermédiaire des concepts et de leurs constituants n'est actuellement disponible.

L'informatique, qui jusqu'alors s'est contentée de faciliter la gestion des données, devrait apporter une véritable contribution portant sur la conception des outils. Les banques deviendront de véritables bases de connaissances, associant différents médias.

Les dictionnaires spécialisés sont bien souvent insuffisants, voire inexistants dans de nombreuses spécialités et pour de nombreuses langues. Quand ils existent, faute

d'une mise à jour fréquente, ils sont inadaptés aux progrès technologiques les plus récents. Une amélioration du contenu des outils et de la méthodologie utilisée pour leur mise au point est à prévoir.

28

Importance de la traduction dans l'aménagement linguistique de la République centrafricaine

Marcel DIKI-KIDIRI

CNRS (LACITO), Paris, France

Résumé

À l'instar de nombreux pays africains, la République centrafricaine vient de s'engager dans la voie d'un bilinguisme d'État en dotant le sängö du statut de langue officielle conjointement avec le français. L'une des nombreuses implications de cette décision est la promotion du sängö comme langue de rédaction des textes officiels (juridiques, administratifs, politiques, etc.). Or une très importante partie de ces textes devra être traduite du français. En outre, l'activité traductionnelle est envisagée comme partie intégrante des techniques d'élaboration des terminologies dont on a besoin dans le cadre de la normalisation et de l'instrumentalisation du sängö. D'où la nécessité de former des traducteurs comme personnes-ressources de l'aménagement linguistique de la République centrafricaine.

La situation sociolinguistique en République centrafricaine

La République centrafricaine compte une cinquantaine de langues vernaculaires parlées régionalement par des groupes ethniques différents. Ces langues sont, dans leur grande majorité, étroitement apparentées. Elles font partie du patrimoine culturel traditionnel et sont parfaitement adaptées à l'expression des besoins langagiers du monde rural traditionnel.

En plus des langues vernaculaires, s'est développée une langue véhiculaire, le sängö, qui s'est répandue sur toute l'étendue du territoire centrafricain, en débordant par endroits sur les pays voisins. Cette expansion géographique s'est accompagnée d'une expansion sociologique, qui a porté l'usage du sängö dans des domaines de plus en plus variés. Parmi les jeunes générations, nombreux sont les locuteurs citadins dont le sängö est la langue première, quelle que soit l'ethnie à laquelle ils appartiennent par leur naissance. Par rapport aux langues vernaculaires, le sängö est incontestablement vu par les populations comme une langue de progrès social, ouvrant sur la vie moderne.

Politiquement, le sängö a toujours été perçu par les autorités centrafricaines comme la langue symbole de l'unité nationale et de l'identité culturelle centrafricaine par-delà les groupes ethniques caractérisés par les langues vernaculaires. C'est pourquoi le sängö a été doté du statut de langue *nationale* en 1964, soit quatre années après l'indépendance, tandis que le français était dès ce moment-là la seule langue officielle. Depuis le 9 février 1991, le sängö a acquis le statut de langue *officielle* aux côtés du français.

Héritage de la colonisation, le français est la langue la mieux équipée pour l'expression de toutes les activités modernes de type occidental : organisations nationales et internationales, administration, commerces, industries, connaissances scientifiques et technologiques avancées, etc. Le français, non seulement domine, mais est perçu comme indispensable au monde de l'écrit. Dans ce domaine, la suprématie du français n'est pas a démontrer. C'est pourquoi le français est la langue de travail de l'État, des sociétés publiques et privées. Or, on estime à moins de 10 pour cent de la population le nombre total des personnes scolarisées capables de maîtriser correctement l'usage du français. La situation est donc telle que 90 pour cent de la population est administrée, jugée, soignée et enseignée dans une langue qu'elle ne parle pas, et qui est la langue officielle de l'État.

La traduction dans l'aménagement linguistique centrafricain

En proclamant le sängö langue officielle conjointement au français, les autorités centrafricaines entendent réduire le fossé qui sépare la population centrafricaine majoritairement « sangophone », de l'appareil d'État. De nombreuses transactions orales se font en sängö. On passe automatiquement au français quant il s'agit de mettre par écrit ce que l'on a conclu oralement. Il faudra donc développer le sängö écrit afin qu'il puisse prendre en charge les actes d'écriture du monde moderne.

La traduction et l'aménagement de la langue de travail

Un des principaux objectifs de l'aménagement linguistique centrafricain concerne l'utilisation légale du sängö comme langue de travail de l'État et des sociétés civiles, partout où il y a interaction professionnelle avec le public.

Ceci implique notamment que soient mis à la disposition du public tous les textes de lois, de règlements, de conventions, etc. dans les *deux* langues officielles. Ces

textes existant déjà en français, il faudra les traduire en sängö. Il en est de même pour les textes statutaires, les programmes d'action, les notes d'information, et tous les textes soumis à l'affichage public. Enfin, de très nombreux formulaires, pièces d'identité ou cartes devront être traduits ou repensés en fonction d'une présentation bilingue.

Il faut bien comprendre que si ces textes doivent être rendus disponibles dans les deux langues, le prestataire de service n'est pas soumis à l'obligation d'une traduction systématique de ses services. Il suffit que ceux-ci soient rendus dans la langue officielle la plus à la portée du demandeur de service. Cette disposition peut, du reste, évoluer s'il se révèle nécessaire de légiférer de façon plus précise sur l'utilisation des deux langues officielles. Pour le moment, il ne semble pas que cela soit nécessaire.

La traduction et l'enseignement des langues officielles

L'introduction progressive du sängö dans le système scolaire est un volet essentiel de l'aménagement linguistique. Il est prévu d'intensifier l'enseignement du sängö à l'université et dans les écoles normales afin de former les futurs cadres et enseignants du pays à la maîtrise égale des deux langues officielles. Mais, à l'heure actuelle, la carence des publications littéraires en sängö est immense. Il y a là un effort considérable à faire non seulement pour traduire en sängö tout au moins les meilleurs auteurs centrafricains et africains écrivant en français, mais aussi pour transposer en sängö les très nombreuses productions littéraires orales existant dans les langues vernaculaires. Les concepteurs de manuels scolaires trouveraient là un fonds inestimable tant pour l'enseignement du sängö que du français.

La traduction et la recherche terminologique

Parmi les activités de développement du sängö pour qu'il puisse assumer pleinement son rôle de langue officielle, il y a un important travail de normalisation à accomplir. Bien que des écrits en sängö existent depuis 1911 (Calloc'h)[1], ce n'est qu'en 1984 qu'une orthographe officielle du sängö, basée sur des travaux scientifiques, a été fixée par décret. Des travaux lexicographiques sont en phase d'achèvement, ils proposeront aux utilisateurs des ouvrages de référence fiables et normalisés. Et le développement des lexiques spécialisés ne fait que commencer.

Lorsque nous parlons de traduction comme partie intégrante de l'activité terminologique, nous devons préciser notre pensée afin d'éviter tout malentendu. La technique qui consiste à prendre une liste de mots hors contexte et à inscrire en face de

1. Calloc'h J. *Vocabulaire français-sängö et sängö-français de l'Oubangui-Chari, précédé d'un abrégé grammatical*, Librairie Paul Geuthner, Paris, 1911. De très nombreux auteurs ont écrit depuis sur le sängö, mais une bibliographie même signalétique nous conduirait bien vite hors du propos de cet article.

chacun des mots de la liste son ou ses équivalents en sängö est à exclure complètement en ce qui nous concerne, non seulement pour des raisons théoriques développées ailleurs[2] mais aussi parce qu'elle produit des résultats très décevants en sängö, notamment à cause des règles de saturation des mots composés dans cette langue.

Nous avons donc développé une approche basée sur la traduction de textes suivis que nous avons appliquée avec succès dans la confection de trois lexiques spécialisés : un lexique de civisme, un lexique de mathématiques (niveau sixième), et un lexique juridique. Dans le premier cas, nous avons d'abord rédigé l'ouvrage en sängö, puis nous l'avons traduit en français. Dans les deux autres cas, les textes originaux étaient en français et nous les avons traduits en sängö.

Au cours de ces traductions, tout le vocabulaire spécialisé est traité en contexte pour garantir la clarté du texte et s'assurer de la meilleure façon de rendre les concepts. La terminologie ainsi produite dans un contexte donné est ensuite vérifiée dans tous les contextes où le terme apparaît afin de s'assurer de la cohérence de ses emplois. Enfin, tous les termes du domaines sont extraits avec leurs contextes d'emploi et placés dans une liste où ils sont comparés les uns aux autres afin de s'assurer que les termes proposés sont les mieux choisis pour les concepts qu'ils doivent rendre, et de procéder, le cas échéant, à des ajustements, voire des harmonisations qui tiennent compte de la totalité des concepts du domaine que l'on s'est donné[3]. Si l'on aboutit à une liste de mots, celle-ci ne contient que des mots dont on sait déjà comment ils se comportent en discours, et pour lesquels des exemples d'emploi sont automatiquement fournis par les divers contextes déjà vus. Et ce n'est pas du tout la même chose que dans le cas où l'on traduirait d'emblée des mots isolés.

La formation des traducteurs

Dans le cadre de l'aménagement linguistique centrafricain, l'Institut de linguistique appliquée a inscrit dans son programme la formation initiale des traducteurs, des terminologues et des assistants d'enquête sociolinguistiques. Cette formation sera sanctionnée par un diplôme universitaire de technologie (DUT). Il est souhaitable que les meilleurs des étudiants issus de cette formation puissent bénéficier de bourses d'étude pour poursuivre et approfondir leur spécialisation dans des institutions du Nord.

2. *Cf.* Halaoui Nazam « Questions de méthode en terminologie des langues africaines », *Terminologie nouvelle*, n° 3 Rint, numéro spécial juin 1990.
3. Cette approche traductionnelle de la recherche terminologique a alimenté une réflexion commune des linguistes africanistes et africains engagés dans la réalisation de lexiques spécialisés dans le cadre du module africain du Rint (Réseau international de néologie et terminologie). *Cf.* note 2.

29

Jargonautes et terminophages ou le traducteur, spécialiste de la langue, au service du consommateur, handicapé linguistique

Louis Yvon CHABALLE et Jean KLEIN

Chaballe Traduction & Communication; Université de Mons-Hainaut, Mons, Belgique

Prologue

« *Der Balluff ist da* ». Traduction française : « Le balluff est là. » Cela ne signifie pas qu'il est présent ou qu'il existe, car même quand il n'est pas là, il est présent.

Par cet exemple, nous voulons montrer ce que peut être le jargon professionnel ou, dans le cas présent, le jargon d'atelier. Un « balluff » est un produit de la firme Balluff de Neuhausen en République fédérale d'Allemagne. Il s'agit d'un « capteur inductif monté dans un portique de chargement d'une cellule de fabrication ».

Le fait qu'il soit là signifie simplement que la DEL (diode électroluminescente) est allumée, ce qui indique que la pince du robot a atteint la butée de fin de course.

Une traduction littérale, comme celle proposée au début, n'est d'aucun secours dans un atelier francophone. Notre propos va tenter d'aborder la manière dont le traducteur peut, dans certains cas — en apportant sa pierre (de Rosette) — éclairer un peu la lanterne du lecteur-consommateur pour lui éviter les affres engendrées par les hiéroglyphes des jargonautes et autres terminophages.

Introduction

Dans un recueil consacré, en partie, à la terminologie, il est inutile, voire peut-être dangereux, de tenter une ou plusieurs définitions des termes « jargon » ou « termi-

nologie », d'abord parce qu'elles sont connues et ensuite parce que ce n'est pas le lieu d'entamer une discussion ou une analyse à ce propos. Contentons-nous donc de renvoyer aux nombreux auteurs qui y ont consacré une partie de leurs recherches ainsi qu'aux normes en la matière, par exemple la norme DIN 2342, partie 1.

L'absence de définitions ne nous empêche cependant pas de recourir à un exemple. Il est emprunté à l'informatique. Comme nous le savons tous, on fait la distinction, dans ce domaine, entre « matériel » et « logiciel ». Dans le jargon professionnel ces termes deviennent respectivement « quincaillerie » et « soft ».

À partir de cet exemple, nous pouvons dégager une tendance générale. La terminologie française de l'informatique privilégie des termes français tandis que, dans le jargon professionnel du même domaine, apparaissent souvent des termes anglais. Cette tendance se manifeste dans tous les secteurs où les Anglo-Saxons, mais aussi les Japonais, jouent un rôle dominant comme dans l'électronique grand public.

Une seconde tendance peut être soulignée : le jargon professionnel — du moins en français — est plus fréquent en langue orale qu'en langue écrite, qui est davantage le domaine de la terminologie. Ceci est peut-être la conséquence du décret paru au début des années soixante-dix interdisant l'emploi de termes étrangers dans les textes officiels et publicitaires français.

Qui utilise la terminologie ou le jargon? Nous envisagerons à la fois l'aspect producteur et consommateur.

Les auteurs de textes

En gros, trois catégories d'auteurs recourent à la terminologie ou au jargon professionnel.

Les scientifiques

Ils peuvent s'adresser à cinq groupes cibles différents. Premièrement, à leurs collègues spécialistes quand, par exemple, ils publient un article dans une revue scientifique. Deuxièmement, aux étudiants en cours de formation dans le domaine lorsque ces scientifiques rédigent des documents pour leurs cours ou des polycopiés destinés à une diffusion interne. Troisièmement, aux techniciens quand ils décrivent un nouveau procédé, une nouvelle machine et son mode d'emploi (notons que ce travail est également réalisé en partie par des rédacteurs techniques). Quatrièmement, aux intellectuels non spécialistes de la matière quand ils écrivent un ouvrage de vulgarisation scientifique. Cinquièmement, au grand public quand ils rédigent un article pour un journal ou répondent à une interview radiophonique ou télévisée.

Ces cinq groupes cibles différents déterminent évidemment le contenu et la terminologie des textes dont la spécificité et le degré de difficulté seront échelonnés en fonction du groupe.

Les rédacteurs techniques

Les groupes cibles auxquels ils s'adressent sont moins nombreux. Il y en a essentiellement trois : les techniciens, les ouvriers qualifiés et le grand public. Ici aussi, le contenu et la terminologie sont déterminés par ces différents groupes. Un exemple éclairant à ce propos figure dans l'ouvrage de Arntz et Picht (1989). Il s'agit du mot allemant *Spiel* désignant le *jeu* entre les pièces mécaniques. Ce mot, défini dans la norme DIN 7182 partie 1 et relevant du domaine de la construction mécanique, devient dans le jargon d'atelier *Luft* et pour le profane *Spielraum*.

Les journalistes scientifiques

Ils ne s'adressent qu'à un seul groupe cible : le grand public. Il s'agit ici de la forme de communication qui est la moins spécialisée, et pour cause. Cela ne signifie cependant pas qu'on n'y a pas recours à la terminologie du domaine, bien au contraire. Toutefois, on n'y retrouve en général que les termes essentiels, indispensables à la présentation de la matière. De tels textes sont hélas souvent des *monstres* parce que dénudés d'isotopie stylistique et terminologique. Il s'agit visiblement du prix à payer pour mettre la science à la portée de tous.

Conséquence

Le contenu et la terminologie d'un texte en langue de spécialité doivent ou devraient permettre une relation harmonieuse entre auteur et consommateur. Si ce n'est pas le cas, ce dernier sera exposé au terrorisme terminologique ou « jargonique », ce qui est loin d'être exceptionnel, car il n'est pas rare que les auteurs perdent de vue le public auquel ils s'adressent.

Les abréviations et les sigles sont un des secteurs privilégiés où s'exerce ce terrorisme linguistique. La situation s'aggrave constamment. Il ne se passe pratiquement pas un seul jour sans que n'apparaissent des néologismes en la matière. Ils sont de tous ordres : officiels, officieux, propres à un seul pays (alors que la même langue est parlée dans plusieurs autres), à une entreprise, voire à un individu. Ainsi, l'abréviation toute personnelle d'un rédacteur technique, pour faire vite et ne pas se donner le mal de recopier chaque fois l'ensemble, est reprise telle quelle par la secrétaire et apparaît aussi telle quelle dans le mode d'emploi sans que personne au monde, hormis le rédacteur, ne sache ce dont il s'agit.

C'est ici que se pose la question fondamentale : que doit faire le traducteur face à ce problème de dysfonctionnement entre auteur et consommateur? Doit-il intervenir, c'est-à-dire doit-il adapter la traduction, devenant par là conseiller au service du consommateur?

Le traducteur au service du consommateur

Définition de la fonction du traducteur en tant que conseiller au service du consommateur

La fonction de conseiller au service du consommateur dépend directement du groupe cible auquel la traduction s'adresse. Elle consiste à mettre la traduction à la portée du groupe cible. Cette intervention facilitatrice pose évidemment un problème épineux : dans quelle mesure le traducteur a-t-il le droit de modifier le texte original pour le rendre plus accessible? Ne court-il pas le risque d'une falsification? Pour éviter cette critique, il faut fixer de manière très précise les limites d'une telle intervention. Elles sont déterminées, d'une part, par les textes eux-mêmes et, d'autre part, par les possibilités d'intervention.

Les textes dans lesquels le traducteur n'est pas autorisé à intervenir

Tous les textes ne peuvent faire l'objet d'une intervention facilitatrice du traducteur. Le meilleur exemple à ce propos est le texte sacré (Bible, Coran, etc.). Saint Jérome, qui était sans conteste orfèvre en la matière, l'avait bien compris et écrivait à ce sujet : « [...] *absque Scripturis Sanctis, ubi et verborum ordo mysterium est non verbum e verbo, sed sensum exprimere de senso.* » Ceci n'a toutefois pas empêché notre saint patron de commettre ça et là quelques erreurs de traduction dans sa célèbre *Vulgate*. Par la suite, ces erreurs ont été institutionnalisées puisque sa traduction a été adopté par le Concile de Trente. Elles se retrouvent d'ailleurs dans toutes les langues. Citons à titre d'exemple la célèbre comparaison affirmant qu'« il est plus facile à un chameau de passer par le chas d'une aiguille qu'à un riche d'entrer dans le royaume des cieux ». Pauvre chameau ou pauvre chas d'aiguille, c'est selon les interprétations. Mais nous serions bien les derniers à critiquer saint Jérome. Que le traducteur qui n'a jamais fauté lui lance la première pierre... Nous connaissons la suite : ... et ils s'en allèrent tous en commençant par les plus vieux.

D'autres textes ne peuvent subir d'intervention facilitatrice du traducteur, par exemple les textes techniques ou scientifiques de haut niveau que le traducteur arrive tout juste à comprendre (ce qui n'est même pas toujours le cas). Ici, toute intervention serait suicidaire. La seule modification possible en la matière est la correction d'erreurs objectives ou matérielles, par exemple lorsque, dans la table des matières, le titre d'un chapitre ne correspond pas à celui du chapitre lui-même ou lorsqu'il y a incohérence entre les numéros de la figure et ceux du texte. Cela arrive même — rarement il est vrai — dans des demandes de brevets et dans des normes. Cependant, en règle générale, le traducteur ne peut ici agir en tant que facilitateur sauf pour les éléments qui, dans le texte, relèvent directement d'un jargon ou d'une terminologie spécifique au pays du texte original. Nous aurons l'occasion d'y revenir.

Après leur adoption, les traités internationaux constituent eux aussi des textes intangibles. Quand des erreurs de traduction s'y sont glissées — mais s'agit-il vraiment

d'erreurs de traduction? —, il n'est plus possible de revenir en arrière. La tristement célèbre résolution 242 des Nations Unies dans laquelle la version française (la libération **des** territoires occupés) ne correspond pas à la version anglaise (*the liberation of occupied territories*) en est un exemple magistral. Ceci a engendré toute la série de problèmes que nous ne connaissons que trop bien puisque les deux versions font foi.

Les textes dans lesquels le traducteur peut ou doit intervenir

Textes où la terminologie et le jargon sont spécifiques au pays du texte original

Il s'agit de textes administratifs, juridiques, économiques (y compris les textes commerciaux et publicitaires) mais aussi scientifiques et techniques. Nombreux sont ces textes qui contiennent — certes, à des degrés divers — des termes spécifiques au pays d'origine. Ces termes sont par définition intraduisibles car, dans la langue d'arrivée, ils n'ont pas de correspondant ni sur le plan du signifiant ni sur celui du signifié. Les exemples abondent et sont connus de tous les traducteurs. Il n'est donc pas nécessaire de les mentionner. Nous nous contenterons de renvoyer le lecteur à nos articles publiés antérieurement.

Textes ou parties de textes inadaptés au groupe cible

Dans les textes destinés au grand public, certains se veulent intellectuels, d'autres scientifiques, etc. L'effet recherché est évident. Les auteurs veulent flatter le lecteur en lui donnant l'impression qu'il est plus intelligent, plus cultivé que ce n'est en réalité le cas. Cela peut fonctionner dans certaines cultures mais pas partout. Citons l'exemple d'un texte publicitaire : une célèbre marque de riz américain a lancé simultanément une campagne publicitaire télévisée sur les chaînes allemandes et françaises. En Allemagne, un noir d'un certain âge, style sérieux et rempli de sagesse, seul devant un paquet de riz explique « scientifiquement » les vertus et qualités du produit. En France, tout était centré sur l'idée de la fête, un repas de noce à l'extérieur, beaucoup de monde, table magnifiquement garnie de mets succulents et du célèbre riz bien sûr avec le même noir, style détendu et bon papa, qui circule en disant « bon appétit mes amis ».

Tout est là!

Autre exemple, qui constitue la réciproque du premier, tiré du secteur automobile : une brochure publicitaire luxueuse de 16 pages en français éditée par une marque d'automobiles japonaises. La traduire telle quelle en allemand constituerait ni plus ni moins une insulte au client potentiel d'outre-Rhin, car cette brochure est littéralement sans véritable contenu. C'est, pour un esprit allemand, un discours vide et superficiel sur le nouveau véhicule. En conséquence, une traduction fidèle est dans ce cas exclue. Il faut récrire un nouveau texte.

Les interventions facilitatrices du traducteur

Le traducteur peut intervenir dans deux domaines, celui de la terminologie et celui de la phraséologie. Quatre types d'intervention sont possibles : la création, l'ajout, la suppression et la modification.

La création

Il arrive que le traducteur, surtout en technologie de pointe où le développement et les réalisations ne suivent pas le même rythme dans les différents pays, doive trouver des néologismes. Plus exactement, il fait différentes propositions parmi lesquelles le client opérera son choix. Parfois, ce dernier les rejette toutes et impose au traducteur une solution, qui aux yeux de ce dernier n'est pas toujours la meilleure, mais... le client est roi.

Certains néologismes peuvent aller de soi et se fondent sur le principe d'identité, comme l'expression *lock-out technique* construite sur le modèle *chômage technique* pour traduire l'allemand *kalte Aussperrung*. Pour d'autres, hélas, cela n'est pas aussi évident. *Cost control* qui devient coûtenance et ses dérivés coûteneur et coûtenir; la maintenique, l'entrepotique, la connectique, la transitique, la documentique, l'interopérabilité, etc.

La création de néologismes par le traducteur doit cependant rester exceptionnelle. Il est exclu que, chaque fois qu'il ne trouve pas un mot dans ses ouvrages de référence ou sa documentation, il recoure à ce procédé. Telle est trop souvent la réaction de jeunes traducteurs encore inexpérimentés qui trouvent par ce biais une échappatoire. Aux réviseurs de réagir contre cette tendance.

L'ajout

Il y a plusieurs formes d'ajouts. Nous n'en examinerons que les principales.

Ajout d'informations complémentaires :

Ils sont nécessaires pour que le client comprenne bien, c'est-à-dire n'interprète pas mal le traduit. Ces informations complémentaires peuvent être présentées sur une feuille séparée quand il s'agit, par exemple, de donner la traduction officielle d'un article de code dont seul le nunéro figure dans l'original. Elles peuvent prendre la forme d'une note du traducteur en bas de page ou en fin de texte pour donner la définition d'un terme usuel dans la langue source mais inconnu en langue cible (par exemple « Conseil général ») ou également pour donner la composition et énumérer les compétences d'un organisme similaire mais différent dans le pays de la langue cible.

Ainsi, il serait hasardeux de se contenter de traduire l'allemand *Betriebsrat* par *comité d'entreprise* ou *conseil d'entreprise* sans autres informations complémentaires, car le lecteur destinataire penserait automatiquement au modèle français ou belge et serait donc trompé.

Ajout d'explications à caractère synonymique ou antonymique :

Dans les modes d'emploi destinés au grand public apparaissent parfois des termes techniques qui ne sont pas susceptibles d'être compris par tout le monde. Dans ce cas, il n'est pas inutile d'expliciter ces termes à l'aide de synonymes ou d'antonymes plus simples. Exemple : les hottes aspirantes domestiques peuvent fonctionner au choix comme appareil avec recyclage ou évacuation extérieure de l'air. Il n'est cependant pas certain que le mot « recyclage » soit compris de tous. C'est pourquoi il est souhaitable que le traducteur ajoute entre parenthèses derrière le mot recyclage l'explication « sans évacuation extérieure » comme le font d'ailleurs de nombreux modes d'emploi rédigés en français.

Ajout d'un mot dans la langue internationale du domaine :

Il s'agit essentiellement du latin et de l'anglais. Dans la traduction allemand-français d'un ouvrage en deux volumes sur les serpents venimeux et non venimeux du continent américain sur lequel vivent plus de 3000 espèces différentes apparaissent évidemment toute une série de noms totalement inconnus sur l'ancien continent. Dans ce cas, le latin est la seule planche de salut possible et, même là, il y a des lacunes. Il n'est d'ailleurs pas nécessaire de changer de continent pour rencontrer de tels problèmes. Juste au-delà de nos frontières, aux Pays-Bas, les Hollandais distinguent quatre types de bourdon pour lesquels ils ont évidemment quatre signifiants différents alors que le français se contente d'un seul.

L'ajout du terme latin est également indispensable pour la traduction de textes de botanique, zoologie, biologie, médecine... ne fût-ce que parce que — contrairement à ce que l'on pourrait croire — les nomenclatures ne sont pas intégralement les mêmes.

L'anglais a également ses domaines où il joue le rôle de langue internationale. C'est le cas de l'informatique par exemple, où l'ajout du terme ou de l'abréviation dans cette langue est fortement conseillé.

La suppression

Suppression de noms propres et de noms de marques de produit :

Dans de nombreux textes apparaissent des noms propres et des noms de marques de produit qui peuvent causer bien du souci aux traducteurs. Dans un article médical en langue allemande sur les prothèses totales de hanche apparaît l'expression *Knochen-Palacos-Komplex*. Après des recherches longues et souvent décourageantes, la solution fut enfin trouvée. *Palacos* est en fait une marque allemande de ciment osseux commercialisé en Allemagne mais parfaitement inconnue en France et en Belgique. Une traduction telle que « l'ensemble os-palacos » aurait donc été parfaitement incompréhensible pour un médecin français, voire francophone. L'omission du nom de la marque allemande s'imposait.

Suppression de dispositions légales spécifiques à un pays :

En dépit des progrès en matière d'harmonisation dans les États membres de la Communauté européenne, il existe encore de nombreuses dispositions légales en matière

de garantie et de sécurité qui ont un caractère strictement national et n'ont pas de correspondant dans les autres pays. Citons l'exemple de l'installation d'une cuve à fioul à l'intérieur du bâtiment. En cette matière, les dispositions légales allemandes sont plus sévères que les belges par exemple. Ainsi, le local où est installée la cuve doit être séparé du reste du bâtiment par une porte blindée, ce qui n'est pas le cas en Belgique. Seul le volume maximal de la cuve est prescrit. Le mode d'emploi allemand pour l'installation de la cuve comporte toute une partie consacrée à la pose de cette porte blindée qui, évidemment, ne doit pas être reprise (traduite) dans un mode d'emploi destiné au marché belge. Certes, il ne s'agit plus ici d'omettre des termes mais bien des passages entiers.

Suppression de mots ou de parties de phrase redondants ou qui pourraient donner au lecteur le sentiment d'être complètement « demeuré » :

Pour les francophones, les langues allemande et anglaise offrent de magnifiques exemples de redondance. Des mots comme *Bohnenkaffee*, *Speiserestaurant* ou *Augenoptiker* sont parfaitement redondants sauf peut-être pour les humoristes qui pourraient distinguer un « opticien des yeux » d'un « opticien-ébéniste » spécialisé dans la fabrication des lunettes de WC. En traduction vers le français, ces termes allemands doivent être obligatoirement amputés de leur moitié.

Par ailleurs, de nombreux modes d'emploi américains présentent une telle redondance et un tel didactisme qu'une traduction fidèle serait le meilleur moyen pour dissuader le client d'acheter l'appareil. Imaginez un texte français du type :

> « Cet appareil possède deux touches. Quand on appuie sur la touche de droite, elle est enfoncée et reste enfoncée. Par rapport à la touche de gauche, elle est plus enfoncée et reste dans cette position tant que vous ne réappuyez pas dessus. Pour la libérer, réappuyer dessus... »

Une seule réaction possible : Au secours, à moi Descartes, Tacite et les autres!

Notons que, dans ce dernier exemple, il ne s'agit plus de terminologie mais de phraséologie, un domaine de la traduction extrêmement important et encore trop peu exploré.

La modification

Modification obligatoire :

Il existe des termes qui doivent être changés en langue d'arrivée. Il a déjà été question des noms de marques de produit qui disparaissaient dans la traduction (*cf. supra*). Dans d'autres cas, le même produit est commercialisé sous des noms différents dans des pays différents. Cela n'est pas rare dans le secteur pharmaceutique. Rappelons-nous le tristement célèbre *softénon* vendu en Allemagne sous le nom de *Contergan*. Le traducteur est confronté à de tels problèmes quand il traduit des ouvrages médicaux dans lesquels des traitements sont décrits ou préconisés. C'est évidemment le nom indigène du médicament qui est utilisé et qui doit être éventuellement modifié dans la traduction pour que le lecteur sache de quoi il s'agit.

Un autre exemple de modification obligatoire sont les mets repris dans les modes d'emploi des cuisinières et des fours. Beaucoup d'habitudes alimentaires, de plats et de recettes de cuisine sont très spécifiques à une région ou à un pays. Ainsi, le plat préféré du chancelier fédéral Helmut Kohl, le *Saumagen*, qu'il n'hésite pas à offrir à ses hôtes de marque, est parfaitement inconnu en France et il est peu probable qu'une « panse de truie farcie » trouve beaucoup d'amateurs dans l'Hexagone, sans parler des *Knödel* et autres étouffe-chrétien d'outre-Rhin.

Ici, il faut obligatoirement transformer ces plats et leur durée de cuisson en les remplaçant par des plats français.

Modification recommandée :

L'adaptation de la traduction au groupe cible et aux habitudes du pays comporte également des changements qui, à première vue, paraissent superflus. Ainsi, dans le mode d'emploi d'une hotte aspirante, l'expression allemande *der gefederte Drehbolzen* pourrait être traduite par « le boulon rotatif monté sur ressort », mais dans un tel mode d'emploi, les francophones préfèrent « système de verrouillage » ou « vis de fixation » non seulement parce que c'est plus simple pour le consommateur mais aussi parce que, dans un tel contexte, les Français préfèrent la fonction de l'objet alors que les Allemands privilégient sa description.

Modification inutile :

Le traducteur pourrait être tenté d'aller trop loin dans son adaptation. Illustrons cela par un exemple. En Allemagne, les prises et les fiches avec terre (*Schukosteckdose* et *Schukostecker*) sont différentes de celles utilisées en France et en Belgique. Les prises ne possèdent pas de broche et les fiches pas de trou correspondant. Les deux contacts assurant la mise à la terre sont disposés latéralement sur la prise et sur la fiche. Lors de la traduction du mode d'emploi d'un appareil électrique avec prise de terre, le traducteur pourrait se montrer perfectionniste et traduire « fiche » ou « prise avec terre, système allemand ». Cette précision est cependant inutile, car les appareils vendus en France et en Belgique auront été modifiés en conséquence, soit par la firme elle-même, soit par l'importateur, et posséderont une fiche avec terre, système français.

Modification impossible :

Parfois il n'est pas possible pour le traducteur de jouer un rôle facilitateur, soit parce qu'il n'est pas capable d'« adapter » un terme ou une expession aux besoins du consommateur, soit parce que son donneur d'ordre le lui interdit. Il n'existe pas que des bonnes relations entre traducteur et donneur d'ordre ou traducteur et chef de service. Les relations peuvent même être parfois tendues.

Exemple : traduction néerlandais-français d'un texte publicitaire pour un appareil de photo reflex. Demande expresse du client : faire court, dimensions maximales 3 × 2 cm.

Traduction proposée : Ayez le bon reflex. Achetez le...

Le donneur d'ordre de langue maternelle flamande a refusé cette traduction, même après les explications du traducteur sur le jeu de mots « reflex/réflexe », considérant qu'il s'agissait purement et simplement d'une faute d'orthographe.

Conclusion

Dans le domaine de la terminologie et du jargon professionnel, il est évident que le traducteur peut et doit intervenir dans un sens facilitateur dans l'intérêt du lecteur. Cependant, ses interventions ne pourront être que limitées et ciblées. De plus, dans ce travail, il sera toujours un peu placé entre l'enclume et le marteau, c'est-à-dire entre le donneur d'ordre et le consommateur. Comme le savent tous les traducteurs, ce n'est pas une position agréable mais, en introduisant dans son opération traduisante cette dimension facilitatrice, en devenant par son action intelligente et efficace traducteur-conseiller, il pourra rendre un service non négligeable à la fois à son client et aux destinataires de la traduction.

30

La station de travail du traducteur professionnel et ses implications pédagogiques

Jean KLEIN et Louis CHABALLE

Université de Mons-Hainaut; Chaballe Traduction & Communication, Mons, Belgique

Remarques préliminaires

Le point de vue développé ci-dessous est celui du traducteur professionnel indépendant, non asservi à un système de traduction automatique, ce qui ne signifie pas qu'il ne puisse y avoir recours de manière ponctuelle.

Le produit envisagé est la traduction finalisée.

Introduction

Pendant les deux dernières décennies, la station de travail du traducteur professionnel a plus changé qu'au cours de toute l'histoire de la traduction. Ceci est dû essentiellement à une révolution :

- **économique** d'adord, ce qui a engendré une circulation planétaire des personnes, des biens et des services;
- **technique** ensuite avec l'apparition d'une multitude de produits et d'appareils destinés à l'usage quotidien mais aussi à la mise en œuvre de techniques des plus sophistiquées;
- **technologique** enfin avec l'avènement de l'électronique et de toutes ses applications.

Cette révolution a modifié fondamentalement l'environnement traductionnel car elle influence de façon décisive les trois étapes de la traduction, à savoir la phase préparatoire, l'opération traduisante et le produit fini, c'est-à-dire la présentation du « traduit ».

Matériels, logiciels et produits informatiques utilisés en traduction

Matériels

Configuration de base actuelle : PC 80286 avec un lecteur de 3,5 pouces (720 Ko), un lecteur 5,25 pouces (360 Ko), un disque dur d'une quarantaine de mégaoctets, une MEV de 1 Mo et une imprimante laser bas de gamme. En deçà de cette configuration, il est pratiquement devenu impossible de faire de la traduction professionnelle.

Configuration idéale actuelle (permettant la PAO) : PC 803886 avec un lecteur de 3,5 pouces (720 Ko ou 1,44 Mo), un lecteur de 5,25 pouces (360 Ko et 1,2 Mo), 4 Mo de mémoire vive, un écran couleur haute définition pleine page (ou mieux encore double page = DIN A3), une imprimante laser avec 20 à 30 polices de caractères résidentes (en EPROM ou sur cartouche), un scanner, un lecteur CD-ROM, disques durs amovibles et/ou un streamer pour l'archivage des fichiers.

Configuration de la station du traducteur en 2001 (voir schéma) : ordinateur personnel à base de 80486, 8 à 16 Mo de MEV, avec lecteur de disquettes haute capacité, disque compact réinscriptible à accès rapide, scanner (OCR et graphique), carte télécopieur/modem sur RNIS, imprimante laser haute définition (1200/2400 ppp), liaison avec des banques documentaires (documentation, terminologie), courrier électronique, impression à distance, liaison avec un système de traduction automatique (TGV = traducteur à grande vitesse).

Grande révolution technologique attendue par tous les traducteurs en l'an 2... : la saisie vocale. Elle présente, par rapport à la traduction automatique, l'avantage de fournir un premier jet raisonné, qui est déjà un premier aboutissement, c'est-à-dire un texte dont les incohérences directes et erreurs orthographiques et stylistiques auront pu être immédiatement corrigées en intervenant sur le clavier. La qualité de ce premier jet permettra une révision efficace, ce qui n'est pas toujours le cas en traduction automatique. En effet, au delà de 25 pour cent de correction, toute révision peut être qualifiée de non rentable. Cela signifie qu'il vaut mieux alors traduire une nouvelle fois le texte plutôt que de le corriger.

L'avantage de la saisie vocale par rapport à la traduction au dictaphone est tout aussi évidente : il s'agit de la vision directe du texte sur écran que ne permet évidemment pas le dictaphone. De plus, dans le cas de phrases difficiles, alambiquées ou particulièrement longues qui ne peuvent être traduites au dictaphone, il est possible de les saisir manuellement au clavier avant de reprendre la dictée.

Logiciels

- logiciels d'exploitation (DOS 5.0 et supérieur, Unix, OS/2...);
- interface utilisateur graphique, orientée objet (Windows...);
- logiciels de traitement de textes (Word [for Windows}, WordPerfect [for Windows], etc.);
- logiciels de conversion (si ces fonctions ne sont pas intégrées dans les nouveaux logiciels; par exemple Software Bridge);
- logiciels utilitaires permettant de restaurer des fichiers endommagés (PC-Tools, Norton Utilities, etc.);
- logiciels permettant la création de terminographies informatisées résidentes ou non en fonction des programmes utilisés (INK, TERM-PC, TERMEX, etc.);
- logiciels de présentation, de mise en page, notamment pour la traduction page à page, les tableaux (Pagemaker, Ventura, Interleaf, etc.);
- logiciels de traduction automatique (Logos, Metal, Systran, Alps, etc.);
- logiciels d'archivage actif de la documentation et des traductions (système du type Canofile équipé d'un écran, clavier, scanner à grande vitesse, disque compact réinscriptible, imprimante laser).

Produits

- banques de données documentaires (Celex, Perinorm, Credoc, etc.);
- banques de données terminologiques (Eurodicautom, Infoterm, Termium, etc.);
- dictionnaires automatiques unilingues ou multilingues (*Robert électronique*, *Harrap's*, dictionnaire de synonymes, etc.);
- correcteurs orthographiques intégrés aux traitements de texte ou indépendants (permettant une analyse contextuelle du type Hugo, Grammatik avec vérification stylistique);
- programme de comptage des mots/caractères du texte (cette fonction est intégrée dans les traitements de texte de la dernière génération — comptage de caractère, de mots... permettant une adaptation aux différents marchés).

Moyens électroniques de transmission du texte traduit

Transmission par support informatique (généralement sous forme de disquette).

Transmission par télécopieur (attention à la qualité du produit). Utilisation de plus en plus fréquente des cartes intégrées.

Transmission par modem avec probablement à l'avenir possibilité de demande d'impression directe chez le destinataire.

En résumé

Face à ce monde — ne devrait-on pas dire à cette jungle — de matériel, de logiciels et de produits, le traducteur se trouve désarmé et, hélas, très souvent grugé par des « marchands de soupe de tout poil » qui profitent de son ignorance ou de son incompétence en la matière pour lui vendre des appareils, des logiciels et des produits inutiles, sous- ou surdimensionnés, peu performants ou franchement mauvais.

Il en va d'ailleurs de même pour les malheureux clients qui commandent des traductions auprès de bureaux ou de services dont l'incompétence n'a d'égal que la malhonnêteté intellectuelle et pécuniaire. On trouvera en annexe à titre d'exemple deux extraits de textes — l'un technique, l'autre touristique — réalisés à l'aide de programmes de traduction automatique. Ils constituent un véritable scandale mais aussi une insulte au client et à la profession de traducteur tout entière.

Les implications pédagogiques

La formation

Il est clair que le traducteur ou l'étudiant en traduction ne peut rater le train électronique. La « traduction d'avant-hier » (papier, crayon ou stylo) « ou d'hier » (machine à écrire), qui existe encore dans certains milieux, est définitivement dépassée sur le plan professionnel. C'est la raison pour laquelle il est indispensable de former les étudiants à la « traduction d'aujourd'hui », c'est-à-dire aux appareils, logiciels et produits qui leur permettront de travailler sur une configuration de base avec les logiciels et les produits indispensables à ce qu'est devenu le métier de traducteur. Or, dans de nombreux instituts universitaires de traduction, cette formation n'est pas prévue dans le programme des cours. Seules les facultés allemandes sont en train de les introduire plus ou moins systématiquement. De plus, cela suppose des investissements importants que les pouvoirs publics ne sont pas toujours prêts à consentir pour des études classées dans la catégorie sciences humaines où l'équipement technique est le parent pauvre et frise l'indécence par rapport aux équipements accordés aux filières scientifiques. Il faut donc trouver des astuces pour permettre aux étudiants de ne pas se sentir ridicules ou révoltés à cause d'une carence de formation quand ils seront confrontés à la réalité professionnelle. Tout comme on n'imagine pas pouvoir former des interprètes de conférence sans cabines d'interprétation, on ne peut plus, à l'heure actuelle, former des traducteurs sans salle d'ordinateurs où ils peuvent disposer du matériel adéquat, à raison d'une configuration de base pour quatre à cinq étudiants, auquel il faut ajouter une vingtaine de postes plus élaborés pour la confection des travaux plus importants tels que les mémoires de fin d'études, notamment ceux qui concernent la terminographie.

Mais le matériel n'est pas tout. L'initiation aux logiciels d'exploitation, de traitement de texte et de terminologie est tout aussi indispensable. Cela permettra d'éviter les nombreux tâtonnements, les erreurs trop fréquentes dues à de mauvaises mani-

pulations, les pertes de fichiers qui peuvent parfois être dramatiques, etc. De plus, la traduction d'un texte présenté selon les règles de la typographie constitue l'indispensable touche finale de l'apprentissage de la rigueur qui est l'essence même de l'opération traduisante et de son inévitable complément, la révision. Ils sont, hélas, encore nombreux les jeunes diplômés incapables de présenter une traduction de qualité professionnelle, c'est-à-dire une traduction dans laquelle une erreur d'orthographe (y compris la ponctuation) par page ou une erreur typographique (y compris les espaces) par page est déjà une erreur de trop.

C'est à ce prix que l'on pourra conférer à la formation un caractère de professionnalisme devenu aujourd'hui strictement indispensable pour permettre une insertion normale dans le métier.

Le recyclage et la formation continue

Si la formation à la « traduction d'aujourd'hui » est indispensable dans les études de traducteur, elle est tout aussi impérative pour pouvoir préparer la « traduction de demain », car les appareils, les logiciels et les produits informatiques évoluent très vite et, si l'on n'y prend garde, on peut se trouver très vite dépassé. Il faut donc se tenir au courant des évolutions et des innovations en la matière.

Outre la formation technique et juridique indispensable, le traducteur du XXI^{e} siècle doit être au fait de quatre grands domaines :

- l'informatique,
- la traductique/terminotique,
- la documentique,
- la transitique.

L'informatique

Planche de travail du traducteur de demain, l'ordinateur personnel du traducteur sera relié à de multiples réseaux de services en tous genres (documentation, terminologie, courrier électronique, etc.)

La saisie vocale nous semble constituer une des principales innovations de demain. Le traducteur dictera directement à son ordinateur qui procédera simultanément à une première correction orthographique.

La traductique/terminotique

Le traducteur — souvent isolé dans un bureau — le sera de plus en plus puisque les contacts personnels seront remplacés par des liaisons visiophoniques et que, sans être physiquement présents sur place, les documents seront accessibles à (très grande) distance.

Le papier disparaîtra de plus en plus de l'environnement traductionnel et se verra remplacer par les supports informatiques. Finis les encyclopédies et les dictionnaires difficiles à ranger et à manipuler. L'image accessible sur une simple pression de touche deviendra l'interlocuteur privilégié du traducteur.

La documentique

Le paradoxe veut que la disparition du papier s'accompagne d'une accentuation du rôle du traducteur dans le cycle de vie du document écrit. Partant de l'écrit et aboutissant à l'écrit, le traducteur devra apprendre à créer, présenter, modifier et archiver ses propres documents. L'archivage — dernier maillon de la chaîne — déterminera le processus même de création. Et, quand on sait que, en vertu de la nouvelle législation sur la responsabilité du fait des produits défectueux, le traducteur peut être tenu pendant trente ans pour responsable des erreurs de traduction, on voit à quel point cette fonction n'est pas à négliger.

La transitique

La multiplicité des plates-formes matérielles et logicielles impose au traducteur — l'une des plaques tournantes dans l'échange des informations — de suivre pas à pas l'évolution de l'échange électronique des données. Son isolement en tant que travailleur et son intense liaison avec les banques de toutes sortes feront de la transmission des données le cordon ombilical propre à assurer sa survie. Chaudement blotti dans son bureau, la vie de ce géniteur à distance ne tiendra qu'à un fil : la ligne de communication.

Conclusion

La profession de traducteur est en pleine mutation. Les exigences s'accumulent dans le sens où il ne suffit plus aujourd'hui de simplement bien traduire, mais il faut aussi savoir présenter son « traduit » grâce aux moyens électroniques modernes. En outre — il ne faut pas se leurrer —, ces exigences vont encore s'accroître. Un jour, on vendra de moins en moins de traductions au profit de produits finis complets, comme des brochures en différentes langues ou une brochure en plusieurs langues imprimées en quadrichromie par le procédé offset. Cela signifie que, au fil des ans, le traducteur risque de devenir de plus en plus un intégrateur de services. Mais c'est là un autre sujet.

31

En dessous de quel seuil de qualité...

Jean-Marie WAAUB

Université de Mons-Hainaut, Mons, Belgique

Les « Industries de la langue », nées vers 1980, avouent un objectif : « initier une stratégie commerciale » et un enjeu : « maîtriser le processus d'industrialisation de l'information »[1]. Sous-domaine de l'ingénierie linguistique, la traductique signifie pour le métier de traducteur le passage du stade artisanal au stade industriel. Le terme *station de travail du traducteur* annonce, pour l'an 2001, la fonction de maillon dans une chaîne de production, maillon robotisable au même titre qu'un poste informatisé dans une chaîne de montage.

Ainsi, une logique de production de masse se développe : le multilinguisme des institutions internationales, l'essor des bureaux de traduction, la professionnalisation des traducteurs individuels imposent la quantité comme facteur prépondérant des coûts de production et donc de la rentabilité, de la viabilité de ces entreprises.

Discréditer le perfectionnisme, tout en sauvegardant une certaine exigence de qualité permet de considérer comme inévitables différentes formes de compromis entre quantité et qualité, notamment en TA et en TAO[2].

1. E. BIERIN, A. MOULIN, F. PICHAULT, *Les industries de la langue : un marché en devenir*, OWIL, Université de Liège, p. 7.
2. Voir notamment W.J. HUTCHINS, *Machine Translation : Past, Present, Futur*, Ellis Horwood, Ltd, 1986; plus spécialement p. 329-331 et notamment :
« Une traduction de qualité médiocre non révisée peut se justifier, si elle est produite en peu de temps et à bon marché, et si elle ne doit pas être publiée. Si on désire une meilleure qualité, il est indispensable de recourir à une collaboration entre l'homme et la machine » (p. 330, ma traduction).

En passant[3], il me suffira de rappeler que le débat, lancé il y a trente ans par BAR-HILLEL, sur les perspectives d'une traduction entièrement automatique de haute qualité (FAHQT, acronyme en vigueur chez les jargonautes) n'a cessé de rebondir[4].

Aujourd'hui, on trouve sur le marché — on nous les a présentés ici-même — divers produits (banques terminologiques, dictionnaires intelligents, encyclopédies sur CD-ROM, et même des analyseurs sémantiques) dont l'ambition est de relever le défi.

Pour évoquer l'état de la question, et tenter d'y répondre, je me propose de mettre sous la loupe un paragraphe, publié le 22 février 1991 par *le Vif/l'Express*, tel qu'il a été traduit de l'anglais au français par un système de TA et révisé par un post-éditeur[5]. Ce texte appartient typiquement à une production industrielle destinée à une communication de masse; l'homme doit ici accepter d'être relayé par la machine. Toutefois, dans cet environnement de machine-prothèse, une situation paradoxale semble s'installer; alors que le technologisme ambiant prétend substituer à la faillibilité de l'homme la fiabilité de la machine, c'est l'inverse qui se produit : on tolérerait que la machine puisse se tromper dans la mesure où l'homme (lecteur naïf, ou traducteur professionnel, ou expert dans un domaine spécialisé) se révélerait capable de la corriger. Telle est l'hypothèse que je me propose de soumettre à une analyse traductologique.

Voici donc le texte source :

> *The parties decided to divide the programme between "core areas" dealing with the essential components of economic transition, and sectoral development needs supporting balanced growth of the economy consistent both with the reform and with easing the hardships endured by the Polish people. The aid will consist of a mix of technical assistance, training and direct financial contributions. Special attention will be paid to the specific needs of the worst affected regions in Poland.*

Sur la qualité de ce texte, qui peut-être n'est pas un original, je m'abstiendrai de tout commentaire. Mais il fallait le traduire.

Voilà le texte brut (*the raw translation*) produit par l'ordinateur :

> Les partis ont décidé de diviser le programme entre les « secteurs du cœur » traitant des composants essentiels de la transition économique, et le développement sectoriel exige la croissance équilibrée de soutien de l'économie com-

3. Les différents articles sont rassemblés dans BAR-HILLEL, *Language and Information, Selected Essays on their Theory and their Appliaction*, Addison-Wesley, Reading, Mass., 1964, Part III; plus spécialement p. 166-179.
4. Même si le terme « traduction de haute qualité » a été souvent critiqué parce que trop vague, les prérequis de toute traduction de bonne qualité définis par Bar-Hillel (*op.cit.*, p. 212) restaient à l'ordre du jour du Colloque CETIL-EUROTRA de Luxembourg (voir documents de travail sur la composante sémantique, 29 mai 1978). Tout récemment, le Colloque de Maastricht des 4-6 janvier 1990 s'est encore une fois penché sur les questions de sémantique computationnelle (voir Marcel Thelen et Barbara Lewandowska-Tomaszczyk (éds), *Translation and Meaning, Part 1*, Maastricht, Euroterm, 1990 et spécialement p. 1-121). Aujourd'hui comme naguère, il appert que certains problèmes n'ont toujours pas trouvé de solution générale automatisable.
5. Le document illustre un article intitulé « La machine contre Babel » dans lequel Olivier Rogeau présente pour un large public le programme EUROTRA et le système SYSTRAN. C'est ce dernier qui est responsable de la traduction en cause (*Le Vif/L'Express*, 22/2/91, p. 22-26).

patible à la fois avec la réforme et avec le soulagement des difficultés supportées par les personnes polonaises. L'aide consistera en mélange d'assistance technique, formant et dirigera des contributions financières. Une attention spéciale sera prêtée aux besoins spécifiques des plus mauvaises régions affectées en Pologne[5'].

Cet objet de consommation est-il consommable (lisible, interprétable)? Quel brouet, quelle basse cuisine, le client boulimique est-il disposé à digérer? Le lecteur naïf trouvera quelque peu barbares les syntagmes *secteurs du cœur* et *personnes polonaises*, mais un fifrelin de bon sens lui permettra de substituer au second l'expression adéquate : *peuple polonais*; par contre il risque de se perdre en conjectures (cardiaques ou nucléaires) pour interpréter le premier. En effet, le terme du texte source (TS) *core areas* ne peut être élucidé qu'à travers une connaissance paradigmatique (*core = most essential part, cf.* Cobuild) et/ou grâce à une perception des correspondances intratextuelles (c'est-à-dire *the essential components of the economic transition*). Autrement dit, il faut relire le TS et l'interpréter, en quoi un traducteur humain eût été infaillible d'emblée. Constatons toutefois que le réviseur de service a corrigé efficacement par *secteurs centraux*.

Comme professeur de traduction, je ne suis pas satisfait du choix de *composants*, là où *composantes* semble s'imposer. Si le réviseur n'a pas jugé bon d'intervenir, c'est qu'il a trouvé que le terme proposé était interprétable; interprétable, certes, mais inexact. Ne jetons pas la pierre au réviseur : le système automatique aurait pu disposer d'un dictionnaire de transfert un tant soit peu plus intelligent, c'est-à-dire capable de distinguer un contexte économique de contextes chimiques ou électroniques.

Un quatrième barbarisme, syntaxique celui-ci, vient en fin de paragraphe hérisser l'épiderme sensible du lecteur : (a) « les plus mauvaises régions affectées en Pologne » — lecteur qui rétablira comme l'a fait le réviseur : (b) « les régions les plus gravement affectées en Pologne »; encore qu'un traducteur robotisé préférera : (c) « les régions de Pologne les plus gravement affectées ».

Sur ce cas, je voudrait faire deux commentaires. Tout d'abord, une évaluation qualitative de ces trois versions; (a) se situer à la limite du non-sens, (b) au niveau interprétable mais incorrect et (c) au rang supérieur : à la fois correct, authentique, et transparent[6]? Deuxième remarque, d'un point de vue computationnel, j'y vois une double insuffisance syntaxique : le programme d'analyse de l'anglais n'a pas été en

5'. Le journaliste du *Vif* affirme que, même s'il reste « quelques fautes », la traduction automatique SYSTRAN est « la plus au point » (sic, p. 25). Dans sa contribution au Colloque de Maastricht (« Bedeutung and Bedeutungsverwendung im Lichte der maschinellen Übersetzung » *loc. cit.*, p. 79-89) Annemarie Schmid a analysé les aspects contextuels des processus de signification que ne parviennent pas à traiter ni SYSTRAN (pour la TA) ni ALPS (pour la TAO). Je n'ai pas l'intention d'ajouter une pierre si petite soit-elle à ce lynchage. Tout comme Annemarie Schmid, je tiens à mettre en lumière des types de comportement que la dynamique du langage inspire à l'homme, en l'occurrence au traducteur, et qui constituent une barrière que ni la TA ni la TAO ne semblent pas pouvoir franchir (contraction d'après le dernier paragraphe, p. 89).

6. Dans les milieux de l'ingénierie linguistique, de telles distinctions sont qualifiées de stylistiques, alors qu'il s'agit de non-sens et de grammaticalité.

mesure d'établir la différence entre *worst* (adv.) et *worst* (adj.); la synthèse du français n'a pas été capable de produire l'ordre des mots adéquats. Une règle contextuelle de transfert pourrait être formulée comme suit :

the worst = V-**ed** + N **le** + N **le plus gravement** + V-**é**

toutefois, l'ambiguïté localisée sur ***worst*** est aggravée parce qu'elle se compose avec l'ambiguïté localisée sur ***affected***; il peut en effet s'agir d'un V-**ed** ou d'un **adj**. On vérifiera aisément que l'**adjectif** possède des traits sémantiques non compatibles avec *regions*.[7]

J'aborderai maintenant deux points cruciaux où le réviseur est intervenu parce que la machine avançait des traductions incorrectes à un niveau élémentaire de syntaxe. Dans chaque cas, son intervention est indispensable car le lecteur naïf ne peut interpréter correctement le produit brut qui lui est proposé. Prenons tout d'abord :

(d) « la croissance équilibrée de soutien de l'économie » qui devient — cela va de soi (cela ne va-t-il pas?) :

(e) « la croissance équilibrée du soutien à l'économie »

avec une restitution plausible de prépositions qui rend intelligible ce qui ne l'était pas. Hélas, trois fois, le beau sens et la belle forme ainsi reconstruits induisent le lecteur en erreur : si nous remontons à la source (il le faut bien...) nous lisons :

(f) *supporting balanced growth of the economy*, c'est-à-dire :

(g) « le soutien d'une croissance équilibrée de l'économie »

— ce qui est une autre histoire.[8] Certes le rédacteur du TS aurait pu trouver meilleure formulation pour sa pensée.

Laissons provisoirement la discussion de ce passage en suspens, pour aborder le deuxième cas où le réviseur a dû intervenir à un niveau de structuration syntaxique plus global. Lorsque le lecteur aborde la troisième phrase,

> « L'aide consistera en mélange d'assistance technique, formant et dirigera des contributions financières »

il se doute — si naïf soit-il — que la machine ne tourne pas rond; que faire de ce « **formant** » totalement déconnecté du reste de la phrase. Ce dont il ne se doute pas, c'est que le programme — qui semble assez intelligent pour coordonner deux membres de phrase au futur (« l'aide consistera... et dirigera... ») — organise en réalité un contresens parfait, notamment parce qu'il ne perçoit pas le polysémie de ***direct*** traité comme verbe et non comme adjectif. Pour établir l'interprétation correcte, le réviseur ne peut s'appuyer sur le simple bon sens : il doit relire le TS de très près et repérer une structure de coordination reliant trois termes, à savoir : « un mélange de A, B,

---

7. Ma remarque n'a pas pour objet de dénigrer le système. Ce qui m'importe, c'est de suggérer deux choses : (a) on se contente de procédures inadéquates pour ne pas alourdir le système et (b) on n'est pas en mesure d'intégrer le savoir-faire du réviseur dans une machine qui ne dispose pas d'un module d'apprentissage.
8. Autrement dit, (d) est une forme agrammaticale, interprétable, tandis que (e) est une forme grammaticale, interprétable aussi; mais toutes deux véhiculent un faux sens, presque un non-sens. La portée de l'erreur reste limitée, il est vrai, mais vous ne perdez rien pour attendre.

et C », où B (***training***) donne **formation** et C (***direct financial contributions***) **des contributions financières directes**. En dérapant sur le domaine de la coordination (**consistera... et dirigera...**) la machine a produit un leurre de carton, contre lequel l'analyse du traducteur intervient avec la rapidité et la précision d'un missile.

Maintenant, revenons à cet autre leurre de carton que même le réviseur humain n'a pas pu débusquer dans la syntaxe de la première phrase. Ce n'est pas par hasard qu'il s'agit une fois encore — en partie du moins — de définir correctement le domaine d'une coordination (***and***). Certes, coordonner, comme le fait la machine, « **les parti(e)s**[9] **ont décidé de diviser le programme... et le développement sectoriel exige...** » est grammaticalement correct. Mais ce texte est-il interprétable? Quelle cohésion discursive trouver à cette coordination de deux assertions dont les propos (thèmes, sujets) ne présentent guère de progression ou d'articulation logico-sémantique et dont les prédicats (rhèmes, verbes) sont actualisés dans des perspectives temporelles et aspectives si peu compatibles?

Par la volonté du rédacteur — celui-ci n'a-t-il pas le droit, la liberté de choisir sa forme d'expression? — nous nous trouvons confrontés à un nœud complexe d'ambiguïtés : domaines de la préposition ***between*** et de la coordination ***and***, ambiguïté morphogrammaticale de ***needs*** (V ou N), et ambiguïté syntaxique de ***supporting*** (COD du V ***needs*** ou expansion du N ***needs***). Il n'empêche qu'un lecteur humain quelque peu sensible aux structures parallèles et entraîné à pratiquer un découpage intelligent lira :

« divide the programme between
(a) « core areas »
dealing with the essential components...
and
(b) sectoral development needs
supporting balanced growth of the economy... »

Ne sommes-nous pas peu fiers d'avoir déjoué les pièges tendus par la machine et par le rédacteur? Or la stratégie adoptée par le traducteur humain ne semble pas facile à définir. En l'occurrence, tout se passe comme si la possibilité de construire mentalement (***with the mind's eye***) un parallélisme structural caractérisé par deux arches de très longue portée avait conduit le lecteur à éliminer toute autre interprétation.

Certes, il convient de se garder de tout triomphalisme humaniste et de vérifier expérimentalement le comportement d'un THNAO — cet éternuement jargonautique désignant le traducteur humain non assisté par ordinateur. J'ai donc soumis ce désormais célèbre paragraphe à la sagacité d'une centaine d'étudiants de première et seconde licences. Les résultats détaillés — repris dans le tableau ci-joint (appendice II) — confirment, si besoin en était, la triple ambiguïté de la première phrase : les sujets se répartissent globalement en trois groupes, en fonction de trois interprétations :

(A) 32 sujets perçoivent correctement le paralléisme; 17 d'entre eux l'explicitent; pour eux ***needs*** est un nom;

9. Le réviseur est bien inspiré de rectifier **partis** en **parties**. Il lève ainsi une ambiguïté purement lexicale. Pour ce faire, il ne peut utiliser aucun élément formel du contexte; il a bien dû faire appel à sa connaissance des événements récents. Quel logiciel pourrait le suivre dans cette voie?

(B) 31 sujets se trompent sur la portée de la conjonction de coordination (dans le domaine de la préposition ***with*** au lieu de celui de la préposition ***between***), tout en interprétant correctement ***needs*** comme nom;

(C) 29 sujets (le petit tiers) se trompent à la fois sur le domaine de la préposition ***between*** (interrompu avant le ***and***) et sur la nature de ***needs*** (interprété comme verbe).

Remarquons que sept sujets ont produit un texte acceptable mais qui conserve la triple ambiguïté de l'original. Enfin, six sujets donnent une version non interprétable.

Il apparaît maintenant que l'ordinateur et le réviseur — le second sous l'influence *maléfique* du premier — se rangent dans le troisième groupe. L'ordinateur est bien excusable d'ailleurs : je ne vois pas qu'une stratégie préférentielle, fondée sur le principe de ***late closure*** ou sur celui de ***minimal attachment***[10], permettrait computationnellement la stratégie des arches parallèles et de portée maximale. Le pourrait-on, il faudrait encore que la machine puisse faire le bon choix entre ces stratégies contradictoires : je ne vois pas sur quels éléments formels du contexte elle pourrait se fonder. Si seulement elle eût pu reconnaître en ***needs*** un nom plutôt qu'un verbe... Comme je l'ai montré plus haut, seule une analyse fine de la structure et de la cohésion discursives l'eût permis.

S'agit-il vraiment de TA ou même de TAO? Je parlerai plutôt, en proférant une syllabe quasi pharaonique, de TMAH, c'est-à-dire de traduction mécanique assistée par l'homme, pour montrer du doigt un système qui fait du traducteur humain un esclave coûteux mais assez fiable, subordonné aux caprices coûteux d'une machine bien trop faillible.

L'environnement majeur que le traducteur humain doit mettre en œuvre reste et restera son propre environnement mental[11]. Pour conclure, je soumettrai à votre réflexion une question : l'explosion quantitative et multilingue justifie-t-elle le recours à la machine-prothèse, si nous sommes menacés de devenir des presque analphabètes consommateurs de textes fétiches, perdus dans un univers de machines qui réduiront le sens au plus petit commun dénominateur de leurs ignorances pour communiquer entre elles et manifester un pouvoir dont nous serons les objets, mais aussi les sujets exclus? Je suis persuadé que cette perspective apocalyptique ne nous menace pas dans un avenir immédiat, mais il appartient aux traducteurs comme aux autres gardiens de la fonction langagière de préserver l'intégrité de celle-ci.

---

10. L. FRAZIER, *On Comprehending Sentences : Syntactic Parsing Strategies*, Ph.D. Thesis, University of Connecticut, 1979, cité par Kurt Haverkort dans sa contribution au Colloque de Maastricht, *loc. cit.*, p. 35.

11. Georges L. BASTIN, « L'adaptation — condition et concept », *Études traductologiques*, 1990, p. 215-230. J'ai quelque peu détourné à mon profit sa remarque :
« Précisons... que le traducteur-adaptateur fait partie du *milieu*. Il est son propre environnement et, de ce fait, impose ou s'impose des conditions par les choix délibérés qu'il fait » (p. 218).

Appendice I

Jan 15 15:06 1991 phare Page 1

The parties decided to divide the programme between "core areas" dealing with the essential components of the economic transition, and sectoral development needs supporting balanced growth of the economy consistent both with the reform and with easing the hardships endured by the Polish people. The aid will consist of a mix of technical assistance, training and direct financial contributions. Special attention will be paid to the specific needs of the worst affected regions in Poland.

Les partis ont décidé de diviser le programme entre les "secteurs du cœur" traitant des composants essentiels de la transition économique, et le développement sectoriel exige la croissance équilibrée de soutien de l'économie compatible à la fois avec la réforme et avec le soulagement des difficultés supportées par les personnes polonaises. L'aide consistera en mélange d'assistance technique, formant et dirigera des contributions financières. Une attention spéciale sera prêtée aux besoins spécifiques des plus mauvaises régions affectées en Pologne.

Appendice II

Interprétations des domaines de *between* et de *and*.

A) **between « core areas »... and sectoral development needs** (= N)

A′) idem + explicitation (d'une part... d'autre part...)

B) **between « core areas » dealing with the essential components of... and/with/ sectoral development needs** (= N)

B′) idem + explicitation (*cf.* A′, ou **avec** X et **avec** Y)

C) **between « core areas »** /./ **and sectoral development needs** (= V)

D) ambiguïté non résolue mais interprétable OK (**needs** = N)

E) interprétation incorrecte ou incohérente

Successivement (dans l'ordre de passation des tests — sur 3 jours)

(a) (b) (c) : 3 groupes de 2^{e} licence; sous-total : t′

(d) (e) : 2 groupe de 1re licence; sous-total : t″

	A	A′	(A + A′)	B	B′	(B + B′)	C	D	E	N
(a)	4	1	(5)	2	1	(3)	(5)	0	2	15
(b)	3	2	(5)	3	1	(4)	(1)	2	1	13
(c)	1	2	(3)	3	0	(3)	(9)	0	1	16
(d)	2	2	(4)	5	0	(5)	(7)	3	1	20
(e)	5	10	(15)	14	2	(16)	(7)	2	1	41
t′)	8	5	(13)	8	2	(10)	(15)	2	4	44
t″)	7	12	(19)	19	2	(21)	(14)	5	2	61
GT)	15	17	(**32**)	27	4	(**31**)	(**29**)	7	6	105

32

L'ordinateur est une vraie machine à écrire

Jean-René LADMIRAL

Université de Paris-X-Nanterre, CERT, ISIT, France

À entendre tout ce qui s'est dit à ce colloque de Mons, j'en étais venu à douter si ma propre communication présentait encore quelque intérêt. Ma contribution est en effet proprement élémentaire — et je dirai, pour parer à une objection qu'on pourrait y faire, archaïque. Et puis une conversation en marge du colloque m'a rassuré, quand l'une de nos collègues nancéenne, très avancée dans la recherche sur l'informatisation de la traduction (qui nous occupe ici), m'a confié (à table) que quand il lui arrivait d'avoir un texte à traduire, occasionnellement, elle en revenait au dispositif crayon-papier. Du coup, il m'est apparu que ce que j'ai à dire sur le sujet avait quand même bien sa place au sein de ce concert de communications souvent très sophistiquées, très « *high-tech* ». Mon propos, en l'occurrence, se situe au niveau « existentiel » de ce que j'appellerais une phénoménologie du vécu de la technologie.

Comme quelques autres, j'entends adopter ici le point de vue du traducteur. C'est sur la base d'une pratique effective de la traduction qu'ont été développées les réflexions qu'on va lire; et surtout elles ne vont qu'à faciliter, à « optimiser » cette même pratique, au niveau minimaliste d'une expérience de la traduction au jour le jour[1]. Concrètement, il s'agit de répondre à un certain type de difficultés que rencontre tout traducteur.

En schématisant, on pourra regrouper les difficultés de traduction sous quelque six rubriques.

1. Sans oublier le niveau didactique de la formation des traducteurs, *cf.* ma communication aux Journées européennes de la traduction professionnelle (UNESCO, Paris, le 25 mars 1987, dont les Actes ont été publiés dans la revue co-éditée par l'Université de Paris-VIII (Vincennes à Saint-Denis) et Hachette : *Encrages*, n° 17, Printemps 1987.

1. On pense d'emblée aux difficultés de langues, et d'abord à celles que présentent la langue source (Lo). Comme s'il suffisait de « maîtriser » une langue étrangère (L2) pour être traducteur!
2. En fait, dans l'exercice de la pratique, c'est plus encore la langue cible (Lt), langue maternelle (L1) elle-même, qui fait problème. Le traducteur ne le sait que trop — et plus encore ceux qui, comme certains d'entre nous, sont en charge de former les apprentis traducteurs...
3. Il y a aussi les difficultés d'ordre terminologique : c'est un point tout à fait essentiel et primordial, notamment dans le domaine, quantitativement privilégié, de la dite traduction professionnelle ou « technique ». L'informatisation apporte là beaucoup à la *station du traducteur de l'an 2001* (ainsi qu'en font foi bien des contributions à notre colloque).
4. De même, il y a les difficultés de traduction qui appellent un travail de documentation, car elles exigent du traducteur une connaissance du domaine (*subject-matter*) : dans le domaine technique, la traduction ne pourra souvent être faite efficacement qu'en faisant fond sur les *realia* auxquels réfère le texte.
5. À quoi on devra ajouter, à mi-chemin des compétences proprement linguistiques et de cette connaissance « sigmatique », une plus ou moins grande familiarité avec l'arrière-plan des présupposés culturels que véhiculent différentiellement les langues en présence, dans la mesure où les traducteurs sont aussi des professionnels de la communication interculturelle[2].
6. Enfin, il y a la difficulté à réexprimer, à reformuler le message en langue cible, après le *saltus* faisant rupture avec les signifiants propres au texte source. Ces difficultés de traduction relèvent de l'écriture, c'est-à-dire de l'*expression* écrite. Ce sont exclusivement ces dernières dont je traiterai ici.

Elles apparaissent dans ce qui constitue pour moi la « seconde phase » de l'opération traduisante. Pour m'en tenir à l'essentiel, je dirai en effet qu'une description adéquate et opérationnelle du fonctionnement de l'opération traduisante en nous permet de distinguer *deux phases*[3]. Quand on consulte les, déjà relativement abondants, ouvrages théoriques sur la traduction, on y trouve souvent des schémas fort complexes et sophistiqués qui, à grand renfort de linguistique et de théorie de la communication, voire de psychologie (de seconde main), s'attachent à pousser très loin dans le détail l'analyse, en multipliant les phases et les instances censées décrire les processus à l'œuvre dans le cours de l'activité traduisante : il en résulte des « algorithmes » de traduction très compliqués, qui d'une certaine façon « font joli », mais ne contribuent

2. C'est un point qu'a évoqué à plusieurs reprises la communication de Karla Déjean Le Féal par exemple. *Cf.* Jean-René LADMIRAL et Edmond Marc LIPIANSKY. *La communication interculturelle*, Paris, Armand Colin, 1989 (Bibliothèque européenne des sciences de l'éducation).
3. *Cf.* Jean-René LADMIRAL, *Traduire : théorèmes pour la traduction*, Paris, Payot, 1979 (Petite Bibliothèque Payot, n° 366), p. 232 et *passim*; ainsi que mon étude sur « Traduction et ambiguïté », *in BRISES* (Bulletin de recherches sur l'information en sciences économiques, humaines et sociales), n° 7, octobre 1985, p. 59 sqq.

que médiocrement à une clarification et à la connaissance réelle des processus effectifs. Je dirai que l'essentiel tient dans les deux phases suivantes :

1. une phase de « lecture-interprétation », où il s'agit de comprendre (de « décoder ») le texte source;
2. une phase de « réécriture » (*rewording*), où il s'agit de produire un texte cible.

Entre les deux, d'une phase à l'autre, il s'opère une « alchimie » psychologique ou psycholinguistique un peu mystérieuse : c'est le *salto mortale* de la « déverbalisation », qu'à notre niveau nous laisserons dans ce que les psychologues appellent une « boîte noire » (*black box*) pour étiqueter leur propre ignorance provisoire. Concrètement, il s'agit de « laisser tomber », d'oublier les signifiants du texte source (To), tout en retenant (aux deux sens du mot) les éléments de signification et même en les intellectualisant, en les conceptualisant, en les faisant passer du verbo-linguistique au logico-cognitif, pour pouvoir les faire réapparaître grâce à un ajustement de signifiants nouveaux dans l'autre langue (Lt). Cette *déverbalisation* va de pair avec une certaine tension psychologique car elle désigne un travail mental qui permet d'assurer l'articulation entre ces deux phases différentes, elles-mêmes matérialisées par des postures corporelles distinctes :

1. une attitude de réception perceptive (et intellectuelle) avec mouvements concomitants des globes oculaires, accommodation du cristallin, etc.;
2. une attitude de production mentale et musculaire, avec subvocalisations phonatoires (?), manipulation (« manuelle ») d'un clavier, ou utilisation (vocale et manuelle) d'un dictaphone, etc.

Curieusement, lesdits ouvrages traductologiques passent en général très rapidement sur cette articulation binaire, problématique mais essentielle, où se fait le travail de la déverbalisation : on se contente le plus souvent d'un étiquetage énigmatique (« transfert », ou même « traduction »...), quand on ne passe pas tout simplement le problème sous silence[4]! Au lieu de cela, on raffine sur l'avant-traduction et sur l'après-traduction, alors que l'essentiel tient, encore une fois, dans ces deux phases dont la déverbalisation assure l'articulation psychologique, binaire. Tout au plus convient-il sans doute d'ajouter une troisième phase — ou plutôt : une « deuxième-et-demie » — qui est la phase des *relectures*, permettant d'assurer une vérification du texte cible obtenu, ce *feed-back* faisant apparaître l'éventuelle nécessité de procéder à certains réajustements ou « réglages » (*cf. inf.*).

Les difficultés de traduction qui m'intéressent ici sont, donc, des difficultés d'expression écrite, que le traducteur rencontre dans la « seconde phase » de son travail. Ce sont des difficultés que je dirai d'ordre à la fois psychologique et littéraire.

---

4. À vrai dire, il existe quand même en traductologie certains chercheurs qui ont commencé à étudier le fonctionnement psycholinguistique du traducteur et de l'interprète : *cf.* Hans P. KRINGS, *Was in den Köpfen von Ubersetzern vorgeht*, Tübingen, G. Narr, 1986 (Tübinger Beiträge zur Linguistik, n° 291); voir aussi les travaux de Erika Diehl (Genève), ainsi que ce qui se fait autour de Danica Seleskovitch à l'ESIT (Paris). Par ailleurs, j'ai publié un premier bilan de mes recherches en psychologie sociale de la traduction dans *La communication interculturelle*, *op. cit.*, p. 21-76.

Sur le plan *psychologique*, ce qui met le traducteur en difficulté, c'est la constante nécessité d'avoir à concilier un ensemble d'impératifs différents, dont l'accumulation fait qu'au bout du compte ils sont contradictoires. Dans l'idéal, en effet, une traduction devra « rendre » le sens exact, mais aussi la connotation, une métaphore ou une « image », telle allusion ou référence culturelle, un jeu de mots reposant sur le double sens possible d'un terme, mais aussi des effets qui se situent au niveau du signifiant, comme des allitérations, un nombre de syllabes (de « pieds »), etc. En soi, pris isolément, chacun de ces aspects est traduisible; mais tout n'est pas « compossible ». Il est *a priori* tout à fait improbable et, dans la pratique, on fait très généralement l'expérience qu'il est effectivement exclu qu'il y ait dans les ressources de la langue cible une équivalence où se retrouvent justement tous les différents aspects qui coïncident dans l'item de la langue source qu'il s'agit de traduire. Il faudrait supposer qu'il y ait entre les langues une sorte d'« harmonie préétablie » panlinguistique, alors que nous faisons constamment l'expérience du contraire en tant que traducteurs (et en tant que linguistes). De fait, certains d'entre nous ont eu l'occasion de vivre des situations d'« ateliers de traducteurs », en groupe, où toutes les solutions proposées à tel problème de traduction sont écartées les unes après les autres, car elles négligent toujours un aspect du texte source; au point que, de proche en proche, la traduction devient impossible.

Ce n'est pas autre chose que nous éprouvons, de façon tacite, dans notre pratique de la traduction individuelle et solitaire. Mais cette accumulation de contraintes contradictoires fonctionne un peu comme le mécanisme du *double bind* mis en évidence par l'École de Palo Alto. Psychologiquement, cette accumulation d'injonctions contradictoires induit une inhibition de l'action, c'est-à-dire de la production des énoncés cible : d'où ces « pannes d'écriture » que nous connaissons bien. C'est pourquoi j'ai pu parler de véritables « blocages » psychologiques et même d'un « complexe du traducteur », avec une perte corrélative des moyens d'expression extrêmement frustrante[5]. Il y a là, pour ce scripteur sous influence qu'est le traducteur, comme l'équivalent d'une « castration » symbolique. Or il m'apparaît que ce type de difficultés de traduction appelle (au moins) deux sortes de réponses. D'abord, à ce problème psychologique, il existe paradoxalement une réponse d'ordre discursif ou intellectuel : une réponse « traductologique », une réponse qui se situe au niveau même du discours théorique *sur* la traduction. C'est une problématique que j'ai développée en d'autres lieux et que je me contente d'indiquer ici[6]. Mais j'avancerai que, non moins paradoxalement, il y a aussi une réponse d'ordre technique, et plus précisément informatique, du côté du traitement de texte.

5. *Cf. Traduire : théorèmes pour la traduction*, *op. cit.*, p. 25.
6. Sur cet aspect de mes « théorèmes pour la traduction », *cf.* mes deux études : « Technique et esthétique de la traduction — Quelle théorie pour la pratique traduisante? », *in* Actes des Journées européennes de la traduction professionnelle (Paris, 25-26 mars 1987), publiés dans la revue *Encrages*, *loc. cit.*, p. 190-197 et « Traductologiques », *in* Marie-José CAPELLE, Francis DEBYSER et Jean-Luc GOESTER (éds.), *Le français dans le monde*, n° spécial : *Retour à la traduction*, août-septembre 1987, p. 18-25.

Psychologique, le problème ainsi posé au traducteur est aussi un problème d'ordre *littéraire*. Non pas seulement au sens élargi et « technique » où les historiens désignent les textes comme des « monuments littéraires » et où il s'agit ici de la traduction des textes *écrits* (par opposition à la « traduction orale » qui est l'affaire de l'*interprète*). J'entends que, dans la seconde phase de son travail, le traducteur se pose des problèmes d'« écriture », au sens plein du mot, où écriture rime avec littérature. Il est dans une situation analogue à celle de l'écrivain confronté au fameux vertige de la « page blanche ». Cela est vrai du traducteur littéraire, bien sûr; mais ce l'est aussi du traducteur « technique », dans une certaine mesure. Certes, les difficultés dont je viens d'allonger la liste comme à plaisir, un peu plus haut, ne se présentent jamais toutes ensemble au traducteur littéraire, ni *a fortiori* au traducteur technique. Il reste que, lorsqu'il s'agit de rédiger un texte cible (Tt), c'est-à-dire d'abord un *texte* tout court, et de triompher des difficultés d'expression que l'on rencontre, la traduction technique elle-même pose aussi *ipso facto* un problème d'écriture, au sens littéraire du terme. En ce sens, entre traduction littéraire, traduction générale et traduction technique, il n'y aurait qu'une différence de degré mais non pas une différence de nature. Il y a une dimension littéraire, fût-elle minimale, dans toute traduction digne de ce nom : les difficultés de traduction littéraires que rencontre la traduction technique sont seulement moins fréquentes et moins délicates.

Le moindre paradoxe n'est pas qu'à ces difficultés de fonctionnement psychologique et de travail littéraire l'ordinateur puisse apporter des éléments de solution. Je ne pense pas ici à la grosse informatique : celle de la dite « machine à traduire » dont, naguère encore, on avait craint qu'elle ne voue la profession de traducteur au chômage. Je ne fais pas référence ici à la TA (traduction automatique), ni même à la TAO (traduction assistée par ordinateur), mais à une utilisation beaucoup plus modeste de l'informatique, à savoir le traitement de texte (TDT), tel que permettent de le pratiquer les micro-ordinateurs personnels (qu'on peut avoir chez soi, comme celui sur lequel je suis en train de « composer » le présent texte). En quelques années, depuis le temps où l'ordinateur inquiétait la profession, les choses ont en effet bien changé et je dirai que maintenant nous avons apprivoisé le « bidule » informatique. Comme le disait plaisamment l'ancienne présidente de l'Association des traducteurs littéraires de France, Françoise Cartano, l'ordinateur est devenu comme un « petit animal familier » qui fait partie de l'environnement de tout traducteur (même littéraire) et dont il ne pourrait plus guère se passer.

On peut résumer *grosso modo* l'histoire des « infrastructures » de l'écriture traduisante en distinguant trois époques :

1) celle de la plume et du système crayon-gomme-stylo;
2) celle de la machine à écrire et
3) celle de l'utilisation des outils informatiques;

c'est-à-dire en l'occurrence, pour ce qui m'intéresse ici, celle du traitement de texte.

On croit souvent que la rupture la plus importante intervient avec l'apparition de l'ordinateur (en passant, donc, de la deuxième à la troisième des trois « époques » qui viennent d'être évoquées). S'il est vrai qu'il en est bien ainsi, dès lors qu'on prend en compte l'ensemble des multiples potentialités techniques que recèle l'ordinateur

mis au service de l'activité traduisante, je préciserai qu'à mes yeux, lorsqu'il s'agit spécifiquement et exclusivement du traitement de texte comme dispositif d'aide à l'écriture traduisante, la rupture se situe ailleurs. C'est quand on passe de la plume (du stylo) à la machine à écrire, du *manuscrit* (MS) au « *tapuscrit* » (TPS), que l'on doit changer sa « technique » d'écriture et de traduction. Alors que le traitement de texte nous permet, pour ainsi dire, de « réinventer le stylo », et le crayon lui-même! En effet, la composition sur écran de ce qui sera un « *compuscrit* » (CPS), sur disquette (puis éventuellement, sur papier), fonctionne un peu comme une sorte de stylo immatériel, mental : comme un de ces « immatériaux » grâce auxquels il devient possible de capitaliser certaines facilités de la technologie moderne tout en revenant aux ressources de souplesse et de créativité propres à l'expérience d'un travail proprement *artisanal* de l'écriture.

Pour en traiter ici, je prends le risque de ne tenir que le discours fragile, léger ou « impressionniste » d'une expérience. Il s'agit de parler de ce que j'ai appelé le vécu de la technologie informatique du traitement de texte : comment vivons-nous la chose? et qu'en faisons-nous? Ma thèse est la suivante : s'il est vrai que, comme le disait (à peu près) Mac Luhan, *the medium is the message*, le paradoxe est que le traitement de texte nous permet, donc, de réinventer le stylo, c'est-à-dire d'en revenir à une écriture traduisante d'avant la machine à écrire. Du même coup, la dimension littéraire ou « rédactionnelle » de la traduction (de toute traduction, encore une fois) ne s'en trouverait pas « aplatie », elle retrouverait là au contraire toute sa place.

Ce qu'en français il est convenu d'appeler une « machine à écrire » (une machine-à-écrire) mériterait mieux de s'appeler une « machine à *taper* ». La vraie « machine à *écrire* », c'est l'ordinateur avec un logiciel de traitement de texte! La dite machine-à-écrire (« sur » laquelle, naguère encore, nous avions coutume de « taper » nos textes) recèle un potentiel d'aliénation, à côté des mérites qu'on lui connaît : elle ne permet guère le retour, les remords d'écriture, les « repentirs »; elle nous oblige à concevoir d'abord entièrement notre phrase avant de la taper, presque à la « subvocaliser » intérieurement, et à la sortir « tout armée » de notre tête. Cela peut convenir à certains écrivains ou scripteurs, mais pas à d'autres : pas à moi notamment, et notamment pas à ces scripteurs-là que sont les traducteurs, quand ils sont confrontés à un texte difficile. C'est à un problème d'écriture — de cette écriture particulière qu'est la traduction, l'« écriture traduisante » — qu'est confronté le traducteur qui est un « co-auteur », un « réécrivain »[7].

Mais si l'on est fondé, dans un premier temps, à souligner les convergences — entre l'écriture proprement dite de l'écrivain et l'écriture traduisante ou « seconde » du traducteur — je voudrais marquer ici une différence essentielle, qui renvoie à une problématique que j'ai thématisée dans les termes d'une *Esthétique de la traduction*[8]. Je n'entends pas par là la célébration « esthétique » de la traduction comme Art de traduire; je prends le mot *esthétique* au sens étymologique, où αισθησις désigne en

7. *Cf. Traduire : théorèmes pour la traduction*, *op.cit.*, p. 22 et *passim*.
8. *Cf.* le numéro spécial, consacré à *La traduction*, de la *Revue d'esthétique*, n° 12 (daté de 1986 — paru en novembre 1987).

grec la sensation, la perception. En ce sens, mon « *esthétique* de la traduction », c'est l'attention portée à ce que j'appellerais volontiers (en pastichant le langage de certains psychosociologues) le « ressenti langagier » de l'écriture traduisante, et particulièrement de la mienne quand je suis en train de traduire. C'est un élément essentiel, coextensif à la réécriture qui constitue la « seconde phase » de l'opération traduisante : c'est le moment du *feed-back* qu'assure le rebalayage du texte cible par une ou plusieurs vagues de relectures. Ce travail se situe dans la foulée immédiate de la réécriture et, en cela, il s'agit bien là moins d'une troisième phase que d'une « seconde-et-demie ».

Quand j'écris « pour mon compte », en « première personne », c'est-à-dire quand j'assiste aux ébats de la petite ménagerie intérieure des idées que j'ai en tête et que je me figure que cela mérite que je l'épingle sur le papier pour en faire un texte, quand je me hasarde à « écrire » donc, j'éprouve très souvent que ce que je viens d'écrire se trouve dire un peu autre chose que ce que j'avais pensé vouloir lui faire dire. Mais très souvent aussi, quand j'ai la surprise de voir s'être glissé sous ma « plume », sous les touches de mon clavier, quelque chose que je n'avais pas pensé dire, il arrive que je me réjouisse de cette espèce de cadeau que vient de me faire mon inconscient (?) et que je décide de l'accepter, de le reconnaître comme s'il était de moi! et d'ailleurs, c'est bien le cas, quand même...

Mais il est bien clair que ce supplément littéraire de l'écriture — ce *débordement heuristique* qu'il lui arrive de nous apporter — que peut (proprement) s'offrir l'écrivain (*stricto sensu*), toute la déontologie du traducteur consistera à se l'interdire. Là où l'auteur peut engranger telle rencontre d'écriture imprévue, presque inconsciente, en se disant : « bonne prise! » — le traducteur dispose quant à lui de cette pierre de touche qu'est le texte original, qui le lui interdit totalement.

Or ce qui est vrai là par excès de l'écriture, au sens restreint, ce l'est aussi par défaut de cette « écriture seconde » qu'est la traduction. Le traducteur devra être vigilant, attentif à ce que son propre texte « veut dire », par lui-même, au-delà de l'intention dont a pu procéder son écriture. Tant il est vrai que nos textes, nos « enfants de papier » acquièrent par rapport à nous la même autonomie que nos enfants de chair, c'est-à-dire qu'ils sont devenus autre chose que nous, une fois que nous leur avons donné naissance. Je suis le premier lecteur de mon texte, non pas seulement au sens où c'est une évidence, une platitude, mais au sens où j'assiste le premier à la façon dont il « fait sens », indépendamment de moi, et éventuellement contre mon intention de départ. Et là, j'indiquerai au passage le risque d'ambiguïté que comporte l'expression de *vouloir-dire*, par ailleurs bienvenue, qu'utilise volontiers Danica Seleskovitch. En effet, il ne faut pas confondre ce que le texte « veut dire » (en anglais : *means*) et ce que son auteur « veut dire » — en deux mots (en allemand : *meint*), au sens d'une intention réelle — ou plutôt : a voulu dire, a *peut-être* voulu dire...

De ce point de vue, si l'on adopte donc l'attitude qui relève de ce que je viens d'appeler une « esthétique de la traduction », le travail de l'écriture traduisante, consistera à procéder à tout un ensemble de *réglages* sur mon texte pour ne lui laisser dire que ce que j'avais voulu expressément lui faire dire. Or la perfection matérielle de la frappe typographique représente une aide déterminante à cette optimisation pro-

gressive du texte cible, car elle permet une *objectivation* de « mon » texte, que je peux lire dès lors comme s'il était d'un autre. À quoi vient s'ajouter l'avantage qu'avec le traitement de texte, c'est d'emblée que j'obtiens cette perfection typographique assurant une lisibilité optimale — laquelle permet au (re-)lecteur que je suis devenu d'« accrocher » tout de suite la moindre scorie d'écriture qui, autrement, aurait pu m'échapper. Surtout, elle se reconstitue perpétuellement sur écran, quelles que puissent être mes corrections; alors que, sur le support papier d'un tapuscrit, c'est très rapidement que les surcharges qu'entraînent lesdites corrections font perdre au texte cette lisibilité transparente sans laquelle je ne puis me mettre dans la disposition « esthétique » de réception, à l'écoute de ce que me dit mon texte, en quelque façon malgré moi.

Plus concrètement, et pour ainsi dire « en amont » de l'analyse qui vient d'être faite : quand le traducteur est « en difficulté(s) », il arrive que ce soit en raison d'une de ces « panne d'écriture », d'un de ces « blocages psychologiques » qu'induit chez lui (au niveau de la « seconde phase » de son travail de traduction, c'est-à-dire au moment de la « réécriture ») la nécessité d'avoir à concilier pratiquement des impératifs dont l'accumulation fait qu'ils sont finalement inconciliables, comme je l'ai montré plus haut; eh bien! il m'est apparu que le traitement de texte apporte, paradoxalement, une réponse informatique à ce problème psychologique. C'est notre expérience commune, à nous autres traducteurs, de nous être trouvés confrontés à des phrases que leur achèvement littéraire rendait « intraduisibles » à nos yeux : que ce soit par excès ou par défaut, qu'il s'agisse d'une phrase décourageante par sa plénitude et son ampleur « proustienne » indominable ou d'une phrase merveilleusement réussie dans le raccourci de son efficacité aphoristique. Il nous est arrivé de nous être laissé fasciner par une telle phrase, d'une façon quasiment « hypnotique ». Tombés en arrêt devant elle, nous restons les yeux fixés sur elle — comme si c'était à l'horizon de cette hypnose, en relisant sans cesse, pour ainsi dire « compulsivement », ces mots du texte source, qu'allait émerger magiquement et l'on ne sait où, comme en filigrane ou entre les lignes de ce texte source, la solution d'une équivalence-cible toute faite. Et pourtant nous savons bien par ailleurs que plus nous sommes dans la dépendance de la fascination, moins nous sommes en mesure de mobiliser nos ressources expressives. Il y a là comme une sorte de castration symbolique, une impuissance d'écriture.

Paradoxalement, dans ce type de situation « nouée », le bidule informatique peut faire des miracles, car il cumule les avantages du stylo *et* de la « machine-à-écrire ». Comme avec mon stylo, je vais pouvoir me hasarder à écrire « n'importe quoi », je veux dire : des bouts de phrases, des esquisses, des variantes, des syntagmes hétéroclites, des éléments fragmentaires d'un *geste phrastique* à venir... et disposer tout cela en attente, dans l'espace non linéaire de la page, crayonner, faire des « graphes », des renvois, etc. — sans m'imposer tout de suite le censure « surmoïque » d'une exigence d'achèvement; alors qu'à la « machine-à-écrire », encore une fois, il faut avoir déjà sa phrase en tête avant de la coucher sur le papier. Bref, je vais pouvoir « débloquer » la situation, *désinhiber* mon écriture, en prenant congé du « perfectionnisme » qui pèse souvent si lourd sur le traducteur. Mais en même temps, comme s'ils étaient tapés à la « machine-à-écrire », ces linéaments d'une écriture encore timide et fragile, inchoative, évanescente, presque vélléitaire, se trouvent d'emblée crédités d'une trans-

parence à la lecture que leur confère la perfection d'une frappe typographique et qui, du coup, induit une dynamique de formulation et de reformulation, permettant ainsi le redémarrage de l'écriture traduisante.

S'agissant précisément de rassembler et de mobiliser nos ressources expressives pour lever le type de difficultés de traduction auquel je fais référence ici et pour produire ainsi un texte cible, nous savons tous l'importance du « premier jet ». Mais il est tout aussi clair qu'il doit être ensuite soumis au contrôle des relectures successives qui permet l'optimisation progressive du texte cible thématisée plus haut. Ainsi entendue, l'esthétique de la traduction a la fonction critique d'un examen par la grâce de ce retour (*feed-back*) du texte cible qui permet de mobiliser la fonction psycholinguistique de contrôle (*monitor*), coextensive à notre compétence linguistique (en L1, voire en L1′, L1″...), en tant qu'elle est en l'occurrence essentiellement une compétence de réception.

Tout mon propos aura été d'indiquer ici comment, même dans des cas difficiles, et particulièrement dans ces cas-là, le traitement de texte permet de mener de pair les deux mouvements opposés et complémentaires dont procède toute écriture, et en particulier l'écriture traduisante : il permet à la fois de cultiver les vertus du « *premier jet* » et de satisfaire aux exigences de l'*achèvement* littéraire.

Mais en l'occurrence, s'agissant spécifiquement de traduction, il aura fallu mettre en place un dispositif de *désurmoïsation* de l'écriture; et je me suis attaché à montrer que le traitement de texte peut fonctionner dans ce sens. Il permet en effet de s'affranchir en partie du « surmoi » linguistique et littéraire relativement tyrannique des francophones (au moins chez les scripteurs de ma génération), qui contribue à rendre l'exercice de la réécriture traduisante si difficile, si pesante. Mais cette désurmoïsation que permet la technologie informatique, par le jeu qu'elle autorise dans cette sorte de jardin d'enfant de l'écriture que nous offre l'écran, ne fait que rendre plus aisée la gestion de la fêlure psychologique du traducteur; elle ne fait qu'induire une certaine désinhibition de l'écriture. Bien évidemment, elle ne dispense pas de la nécessité constante de faire des *choix de traduction*[9] : en elle-même, elle ne règle pas le problème des « variantes ».

Il reste que la possibilité de tripoter, de tapoter un peu « à l'aveuglette » ou « à tâtons » sur les touches du clavier, de « faire joujou » avec le bidule informatique est de nature à nous affranchir des *pesanteurs* de l'écriture traduisante, de cette impression que nous avons parfois que tout cela pèse des tonnes! C'est peu, et c'est beaucoup. Paraphrasant Georges Brassens, je dirai que — comme d'autres choses parfois — l'écriture traduisante, « ça s'commande pas »! enfin, pas toujours...

L'expérience de traduire, c'est qu'il arrive qu'on soit tellement dans la dépendance du texte original, pour ainsi dire sous sa coupe, qu'on va projeter le souvenir du sens qu'on a du texte source sur le texte cible dont on a conçu le brouillon; en sorte qu'on ne sera plus en mesure d'en percevoir, d'en « recevoir » (*rezipieren*), les faiblesses. Bien plus, quand on a un manuscrit (papier), ne fût-ce qu'un tout petit peu surchargé, les difficultés matérielles de relecture (*cf. sup.*) font qu'on doit « accompagner » le

9. *Cf. Traduire : théorèmes pour la traduction*, *op. cit.*, p. 211 sp. et *passim*.

texte ainsi écrit (au brouillon) d'un effort de signification, on est en somme conduit à le « porter » par une subvocalisation intériorisée. On est en quelque sorte dans une attitude d'« assistance maternelle » par rapport à son « enfant de papier »; et cette sollicitude fera encore qu'on n'en verra plus les faiblesses. C'est pourquoi il va falloir l'objectiver et travailler à établir une déconnection, un *découplage* de ces « deux phases » de l'opération traduisante dont il a été question au début de la présente étude. Paradoxalement — à l'instar de l'écrivain, mais *a contrario* — il faudra s'exercer à écouter sa propre traduction comme si c'était un texte original (comme un texte qui n'aurait pas été « traduit-de ») et, donc, comme le texte d'un autre. D'où, encore une fois, l'intérêt stratégique de la perfection matérielle immédiate de la frappe que permet le traitement de texte.

Les problèmes de traduction dont j'aurai traité ici sont donc bien des problèmes d'écriture. C'est pourquoi j'ai parlé d'*écriture traduisante*. La problématique que j'ai indiquée plus haut dans les termes d'une « esthétique de la traduction » en marque la spécificité, puisqu'il s'agit de l'écriture cible d'un sens source qui lui est préexistant. En ce sens, l'écriture traduisante est un cas remarquable, plus encore qu'un simple cas particulier, de l'écriture en général. Cela dit, il n'est pas vrai non plus que le sens à traduire existerait préalablement sur le mode d'une sorte d'objectivité « ontologique ». Il se joue dans le travail de l'écriture (traduisante) une dialectique de l'adéquation réciproque qui, dans les cas difficiles, exige un va-et-vient entre les deux phases de l'opération traduisante. Le sens ultime du texte source « advient » finalement dans l'achèvement de l'écriture du texte cible, tant il est vrai qu'en réalité « on ne traduit pas tant ce qui est écrit, que ce qu'on pense qu'a pu penser celui qui a écrit ce qu'il a écrit quand il l'a écrit[10] ».

Le moindre paradoxe n'est pas que la traduction puisse être et soit si souvent un moyen d'accéder à sa propre écriture (écriture première, majeure). Au reste, on se souviendra du conseil que nous donnait nos anciens maîtres, à savoir de *recopier* des pages de Stendhal ou de Valéry, par exemple, pour apprendre à (bien) écrire. Là, le paradoxe est encore plus éclatant! La traduction est une méta-écriture, une réécriture : traduire, c'est un peu recopier. C'est au demeurant encore le même paradoxe que connaissent bien les philosophes, puisque c'est dans la foulée même de l'apparente soumission au maître que le disciple conquiert l'autonomie de sa raison, de la raison qu'il laisse affleurer en lui, et qu'il en vient à naître à sa propre pensée, à engendrer une philosophie qui lui sera propre.

Dans cette perspective, le recours à l'ordinateur et au traitement de texte n'est pas seulement l'utilisation d'un outil « technique » au sens où il ne ferait que faciliter, alléger et accélérer les choses, sans rien changer d'essentiel. Il y a là en effet deux thèses : pour certains, comme pour Jean-Louis Lebrave[11], l'ordinateur n'est que ça;

10. *Cf.* Jean-René LADMIRAL et Edmond Marc LIPIANSKY, *La communication interculturelle*, *op. cit.*, p. 53.

11. *Cf.* notamment Jean-Louis LEBRAVE, « Le traitement de texte : machine à écrire ou instrument d'écriture », in *Le texte et l'ordinateur* : numéro spécial de la revue *LYNX* (Linguistique Institut Nanterre Paris-X), n° 17 (1987), p. 12-20.

mais on peut soutenir aussi, comme je me suis attaché à le faire, que le traitement de texte induit une dynamique spécifique de l'écriture (traduisante). Encore une fois, ce qui est vrai ici de la traduction l'est aussi de l'écriture tout court, *mutatis mutandis*; et ce que j'ai indiqué ici rejoint les travaux que mènent d'aucuns dans le cadre du dispositif des « ateliers d'écriture »[12]. C'est aussi le sens de mon titre, où n'apparaît pas le concept de traduction, en partie parce que le contexte du colloque de Mons rendait cette information redondante et en partie parce que mon propos aura été de traiter de l'écriture traduisante comme cas remarquable de l'écriture.

À vrai dire, le choix de ce titre résulte aussi des critiques adressées par Christian Boitet à la platitude totale de mon titre initial (« Le traducteur et le traitement de texte »); alors qu'il avait par ailleurs l'amabilité de me féliciter pour ma communication elle-même, où il a vu un produit « haut de gamme » et que, à l'en croire, j'aurais pu me dispenser de qualifier d'« archaïque » (comme je l'ai fait au départ). Si je me permets ces quelques citations pour ainsi dire autophilologiques, c'est qu'elles sont l'occasion de pointer la différence entre la « sophistication » technologique de la plupart des autres communications et ce que j'ai tenté de faire ici.

Mon propos a été de tenir ici un « discours *soft* » : le discours fragile et subjectif d'une phénoménologie, dont je postule qu'elle induit une facilitation psychologique de l'opération qu'elle décrit. L'essentiel étant la dimension de re-*création* du texte. S'il m'est permis de reprendre à mon compte l'idée (flatteuse) que c'est là un produit « haut de gamme », c'est au sens où il faut bien n'y voir que du *soft-ware* intellectuel, pour ainsi dire — par opposition au *hard-ware* informatique, logiciels compris! Je pourrais aussi parler de « bricolage », au sens paradoxalement positif qu'a pu prendre le mot dans le contexte épistémologique des sciences sociales, avec Claude Lévi-Strauss et Pierre Bourdieu. Au delà de cette conclusion, je veux faire encore trois remarques, comme en post-scriptum.

D'abord, il faut bien dire qu'au rebours de tout ce qui vient d'être dit, certains scripteurs préfèrent encore écrire sur du papier plutôt que de concevoir sur écran. Ce sont sans doute en partie ceux qui s'arrangent bien de taper leurs textes directement à la machine à écrire (*cf. sup.*). Il y a là des différences de personnes, et certainement aussi de générations... L'équation personnelle joue un très grand rôle puisque, lorsqu'il s'agit d'écriture traduisante, c'est finalement du rapport éminemment individuel que chacun de nous entretient à l'écriture en général (*überhaupt*) qu'il est question ici; et la phénoménologie que je me suis hasardé à en proposer ne peut manquer de renvoyer de façon déterminante à ma propre subjectivité. C'est ainsi que pour moi — quand je suis traducteur, mais aussi quand j'écris pour mon propre compte, comme c'est le cas avec la présente étude — la composition sur écran *directement*, sans passer par l'étape préparatoire d'un « manuscrit » (écrit à la main), est une aide décisive à la production écrite. D'autres préfèrent en passer (encore) par la phase de l'écriture manuelle; et certains ne se servent du traitement de texte que comme d'une machine à écrire

12. Voir les travaux de Claudette Oriol-Boyer et de l'équipe de Grenoble qui édite la revue *TEM* (texte en main). Dans le contexte qui nous occupe, citon *TEM*, n° 3/4 : « Écriture et ordinateur », Hiver 1984/Printemps 1985.

perfectionnée, pour y taper leurs textes achevés précédemment écrit sur papier. S'il est vrai que le travail du manuscrit est un peu lourd, on conçoit que d'aucuns y aient du goût. Écrire au stylo confère une sorte de présence charnelle au texte : le crissement de la plume contre la feuille, un frottement qui est presque une caresse, et la mouillure de l'encre. Comme en écho à Robert Charlebois, qui « roule à bille sur du papier », je dirai qu'on peut y trouver un plaisir qui fait penser à celui qu'on aurait à skier sur le papier : il y a une surface blanche, immaculée, sur laquelle on laisse la trace mouillée d'une souillure qui est celle de ma main, de mon corps. On peut penser aussi à d'autres analogies plus profondes, plus intimes...

Mais on pourra aussi trouver un inconvénient à l'utilisation du traitement de texte par le traducteur. La possibilité de corriger indéfiniment son texte sur écran diffère d'autant l'échéance *du* texte « définitif ». Tendanciellement, c'est l'idée même que nous nous faisons d'un texte qui est là remise en cause. À plus forte raison pourra-t-on dire de la traduction qu'elle est, comme la psychanalyse, « interminée, interminable » (pour reprendre les termes d'une traduction, elle-même contestable), ce qu'elle était déjà en raison de la nécessité de la remettre sans cesse sur le métier. Le traitement de texte vient là « en remettre » sur ce qu'on pourrait appeler le « complexe de Sisyphe du traducteur », qui n'en a jamais fini de négocier avec lui-même pour amender ses compromis d'écriture, toujours plus ou moins susceptibles d'être améliorés. Il est clair qu'il y a là une aporie, angoissante pour toute écriture, et beaucoup plus encore pour l'écriture traduisante. Mais est-ce que s'en plaindre, ce n'est pas trouver que la mariée est trop belle?

Enfin, il y aurait matière à esquisser une complainte de l'utilisateur de traitement de texte. On ne parle pas assez de ses malheurs, de l'horreur des textes perdus : il y a ses fausses manœuvres, mais il y a aussi les sautes de courant électrique, et les pannes de l'ordinateur lui-même. Il faut rappeler qu'on doit faire des doubles et multiplier les sauvegardes. Plus grave encore : il conviendra de se prémunir contre les « virus ». Et si l'on doit faire réparer son ordinateur, il faudra s'attendre au pire!

33

La formation du traducteur en l'an 2001

Karla Déjean Le Féal

Université Paris 3, Paris, France

Partout dans le monde, la traduction pragmatique fait piètre figure. Une grande partie des traductions est de qualité médiocre, voire franchement mauvaise.

Certes, toutes les traductions ne sont pas l'œuvre de traducteurs professionnels. Certaines sont effectuées sporadiquement par des personnes qui connaissent une langue étrangère et se sentent, de ce fait, une vocation pour ce travail, surtout s'il s'agit de textes relevant de leur spécialité.

Mais bien qu'ils aient alors l'avantage de connaître à fond le sujet traité, le résultat de leurs efforts ne rend que rarement compte de cet atout. C'est qu'ils ne maîtrisent pas la technique de traduction et ne peuvent donc pas tirer profit de leurs connaissances du sujet. Ce sont ces dernières, en revanche, qui font le plus souvent défaut aux traducteurs professionnels, d'autant plus que la technicité des sujets augmente à mesure que la spécialisation devient la règle dans presque tous les domaines.

Toutes les conditions sont donc réunies pour faire échec à la traduction dans bien des cas. Il faut de plus y ajouter ceux où la traduction est effectuée vers la langue étrangère ou en l'absence de connaissances suffisantes de la langue de départ.

La maladie dont souffre la traduction est multifactorielle. La méprise sur la faculté de traduire, assimilée à la connaissance d'une langue étrangère, et la sous-estimation des compétences requises pour la traduction, qui en résulte, en sont tout aussi responsables que l'absence de protection légale du titre ainsi que — et c'est là un point important — l'inadéquation de la formation en traduction.

En effet, le principal mérite de la formation actuelle est de rehausser le niveau des connaissances linguistiques des traducteurs professionnels. Son véritable objectif, par contre, à savoir l'enseignement de la technique de traduction, s'en trouve encore

trop souvent éclipsé. De plus, en associant l'enseignement de la traduction à l'enseignement des langues, elle entretient le malentendu sur la traduction dans le grand public et réduit par conséquent davantage encore son impact sur les données de la pratique. Pour les mêmes raisons, elle attire ceux qui n'ont souvent pas un profil de traducteur et rebute ceux qui l'auraient peut-être. Finalement, bien qu'elle passe pour professionnelle, elle forme, dans l'hypothèse la plus favorable, des techniciens de la traduction, mais non pas de véritables professionnels de la comunication multilingue. De ce fait, elle n'est pas en mesure de faire avancer la professionnalisation de cette activité, qui continue d'être organisée et gérée par des non-spécialistes.

Pour que la traduction puisse relever le défi de notre civilisation moderne, il faut réformer la formation. En effet, sans être le seul facteur de progrès, elle en constitue néanmoins le plus important puisque ce n'est qu'à travers elle que l'évolution du savoir traductologique peut avoir un effet sur la pratique de la traduction, sur sa qualité et donc sur son utilité pour la communication internationale.

Encore faut-il que la formation elle-même — son objectif, son organisation et son contenu — se mette au diapason de cette mission. Il faut donc la réformer.

L'objectif de la formation rénovée

L'objectif en l'an 2001 doit être la formation de véritables professionnels de la communication multilingue. À l'heure actuelle, la solution des problèmes qui se posent dans ce domaine est laissée aux employeurs et aux donneurs d'ouvrage. Or, ceux-ci ignorent le plus souvent les conditions qui doivent être remplies pour le bon fonctionnement de la communication. Le traducteur ne sert généralement que d'outil à l'exécution de desseins qui sont nés avant tout de préoccupations financières et matérielles. Leur efficacité est rarement mise en cause, car les ratés de la communication ne sont pas toujours apparents pour ceux qui en sont responsables et quand ils le sont, c'est à l'exécutant qu'on en tient rigueur et non pas à l'employeur ou au donneur d'ouvrage qui, sans le savoir, est à l'origine d'une commande insensée.

Or, la traduction ne pourra remplir pleinement sa fonction que dans le respect des contraintes et des règles fondamentales de la communication multilingue. Comme c'est le traducteur qui devrait être particulièrement bien placé pour connaître ces contraintes et ces règles, c'est à lui que devrait revenir le rôle de stratège de la communication.

Pourtant, la plupart des traducteurs actuels ne seraient pas capables d'assumer ce rôle, même si on le leur proposait, puisque leur formation, pour autant qu'ils en ont reçu une, ne les y a pas préparés.

Bien au contraire, la formation telle qu'elle est pratiquée aujourd'hui constitue un frein à l'évolution de la fonction du traducteur parce qu'elle cherche non pas à optimiser la communication transculturelle, mais simplement à satisfaire la demande du marché. Ainsi, la plupart des écoles enseignent le thème et attestent à leurs diplômés qu'ils le maîtrisent alors qu'elles savent ou devraient savoir que le thème ne peut jamais répondre aux normes professionnelles de qualité. En occultant ce fait, non seulement elles maintiennent les étudiants dans l'ignorance des règles élémentaires à respecter en

traduction, mais les incitent à les violer, pour arranger l'employeur ou le donneur d'ouvrage qui ne voit que la commodité de la solution sans savoir qu'il sera le premier lésé.

Afin de rompre ce cercle vicieux, il faut tout d'abord changer l'orientation de la formation. Au-delà de l'enseignement de la technique de traduction, elle doit viser une prise de conscience des mécanismes mentaux qui entrent en jeu dans la communication transculturelle, des conditions nécessaires à son fonctionnement, mais aussi des possibilités que les technologies modernes peuvent offrir à la communication multilingue pour peu qu'elles soient utilisées à bon escient.

Pour atteindre cet objectif, il est indispensable de réformer aussi l'organisation et le contenu de la formation.

La réforme de l'organisation de la formation

Cette réforme doit passer obligatoirement par la séparation de la formation en traduction de l'enseignement des langues. En effet, la formation de traducteurs n'a pas davantage à voir avec l'apprentissage d'une langue étrangère que la formation de pilotes, par exemple. Dans les deux cas, des connaissances linguistiques constituent un des préalables à la formation, sans pour autant se confondre avec elle. De même que le pilotage d'un avion s'apprend ailleurs que dans les cours d'anglais, la traduction doit donc s'établir comme une discipline à part entière loin des écoles de langues.

Sortie de l'imbroglio des disciplines et purgée de tous les corps étrangers, la formation en traduction peut facilement être réduite à une durée d'une année universitaire au maximum. En effet, il ne faut pas davantage de temps pour apprendre ce savoir-faire qui n'est que l'orchestration, à des fins de traduction, des capacités et des connaissances qui participent au processus de communication par-delà la barrière des langues.

Une formation de courte durée qui affiche clairement son caractère non linguistique est susceptible d'attirer de nouvelles catégories d'intéressés qui boudent actuellement les cours de traduction : ceux qui connaissent déjà une langue étrangère et ne croient donc pas avoir besoin d'une formation, tout comme ceux qui n'ont pas l'intention de faire de la traduction leur métier et, de ce fait, ne sont pas prêts à investir beaucoup de temps dans une formation supplémentaire.

Chacun d'eux constitue une source potentielle de mauvaises traductions et d'idées reçues. Qu'ils remplissent les conditions de formation ou pas, il y a de fortes chances de voir ces sources se tarir progressivement, car même les candidats recalés seront vraisemblablement amenés à réviser leurs idées sur la traduction et sur leur propres compétences en cette matière. Autrement dit, par le seul fait de sortir de la nébuleuse linguistique, la formation peut contribuer à faire vasciller les préjugés et à éclaircir la situation.

Mais avant tout, il y a fort à parier que parmi ces nouveaux intéressés, on trouvera bon nombre de candidats qui répondront mieux que les étudiants en langues au profil du traducteur de l'an 2001. En effet, les connaissances non linguistiques nécessaires

au traducteur ne peuvent plus dans tous les cas être acquises sur le tas. Elles demanderont de plus en plus souvent des études approfondies du domaine concerné, ce qui suppose des aptitudes et un véritable intérêt pour cette matière. Il y a plus de chances de voir ces conditions réunies chez ceux qui ont spontanément opté pour ces disciplines que chez ceux qui se sont tournés vers les langues.

Il s'y ajoute que les premiers sont davantage susceptibles que les seconds de trouver leur compte dans l'activité traduisante, dont l'attrait est davantage fonction de l'intérêt qu'on porte aux sujets que de celui qu'on porte à la langue étrangère.

Si l'on veut que la formation se limite à ce qui est propre à la traduction, il faut s'assurer que les candidats possèdent les capacités et les connaissances qui entrent en jeu dans le processus de traduction. Ils doivent donc avoir des capacités certaines d'analyse et de synthèse, une bonne maîtrise de leur langue maternelle à l'écrit, les connaissances non linguistiques nécessaires à la traduction dans les domaines où ils comptent travailler et une compréhension suffisante d'une langue étrangère.

Il est indispensable que l'ensemble de ces préalables soient réunis; si l'un d'eux fait défaut, le processus de traduction ne peut s'accomplir normalement. Tous les efforts didactiques seraient alors vains. Ce serait vouloir enseigner la pâtisserie sans farine. Et de même que la cuisine n'est pas l'endroit qu'il faut pour faire pousser le blé, de même le cours de traduction n'a pas pour fonction de créer un bagage cognitif, mais d'apprendre à s'en servir dans un but déterminé.

La réforme du contenu de la formation

Bien que le processus de traduction comporte deux phases, celle de l'analyse du texte et celle de la réexpression, ni l'une ni l'autre n'est spécifique de ce processus. Prises séparément, chacune de ces opérations mentales fait partie de nombreuses autres activités humaines. Que l'une ne se fasse pas dans la même langue que l'autre n'est pas non plus le propre de la traduction. Du moment qu'on peut lire dans une langue étrangère, on s'approprie des connaissances et des idées qu'on est amené, tôt ou tard, à réexprimer dans sa langue maternelle sous une forme ou sous une autre.

Pour qu'on puisse parler de traduction, il faut que ce soit exactement les mêmes idées qui soient réexprimées, et cela dans le même but et sous une forme qui produit le même effet. C'est cet impératif d'équivalence qui constitue le lien entre les deux opérations.

En raison des différences entre les langues elles-mêmes et la manière de les utiliser, on ne peut réaliser cette équivalence qu'en dégageant le sens du texte pour le reformuler ensuite dans l'autre langue.

À première vue, ce passage obligé par le sens ne devrait pas poser problème du moment qu'on dispose des connaissances nécessaires à la compréhension du sens du texte et à l'expression de ce sens dans sa langue maternelle. Mais pour les besoins de la traduction, il faut parvenir à débarrasser complètement le sens des formes linguistiques de départ afin de pouvoir lui imprimer les formes propres à la langue d'arrivée.

C'est cette déverbalisation totale qui est à la fois particulière à la traduction et difficile à réaliser.

Elle ne réussit d'ailleurs que s'il y a simultanément activation des formes adéquates dans la langue d'arrivée. Celles-ci exercent, tels des aimants, une attraction sur le sens qui lui permet de se libérer complètement de la langue de départ et de passer sans encombres dans la langue d'arrivée.

Cette déverbalisation totale du sens et l'activation concomitante des structures d'accueil représentent le savoir-faire propre du traducteur. C'est donc sur ces opérations, qui impliquent une mobilisation maximale des connaissances linguistiques et non linguistiques dans un effort conjugué d'analyse et de synthèse, que doit se focaliser l'enseignement. En effet, les étudiants doivent prendre conscience de la démarche à suivre et en acquérir la maîtrise.

À l'heure actuelle, l'enseignement porte essentiellement sur ce qui précède et sur ce qui suit cette opération charnière du processus de traduction. L'opération elle-même est traitée comme la boîte noire à laquelle on ne touche pas. C'est ce qui explique qu'on s'attache davantage aux préalables qu'au processus même de la traduction et que la formation se termine souvent là où elle devrait commencer.

Par conséquent, les méthodes didactiques doivent être réformées elles aussi. Il faut mettre au point des méthodes qui permettent de guider l'étudiant pas à pas à travers cette phase critique du processus jusqu'à ce qu'il ait suffisamment intériorisé la démarche pour pouvoir la suivre seul et sans aide. Une méthode de ce type est en cours d'élaboration.

Si l'on considère que le passage d'une langue à une autre est un processus dynamique dont la force motrice consiste en un effort conjugué de déverbalisation et d'évocation de formes verbales adéquates au sens dégagé dans la langue d'arrivée, on comprend pourquoi le thème ne peut pas être appris. C'est qu'une langue étrangère n'exerce pas la même attraction que la langue maternelle sur la pensée qui cherche à s'exprimer. Par conséquent, le processus manque de dynamisme. Il s'y ajoute que le peu d'attraction qu'il peut y avoir est souvent exercé par des formes verbales supposées adéquates, alors qu'elles ne le sont pas en réalité.

C'est pour les mêmes raisons que le processus de traduction ne peut pas aboutir lorsque les connaissances non linguistiques sont insuffisantes. Mais alors que ce type de déficit peut le cas échéant être comblé, l'inaptitude de la langue étrangère à la fonction de langue cible est quasiment irrémédiable; si l'on peut apprendre à mieux manier une langue étrangère, on aura toujours plus ou moins de mal à « sentir » cette langue. Or, c'est le sens de la langue qui est à la base du magnétisme propre à la langue maternelle. Comme il ne se développera jamais suffisamment dans la langue étrangère, celle-ci est, par son essence même, inapte à servir de langue d'arrivée. Il suffit de lire n'importe quel thème pour s'en convaincre.

Le thème n'aura donc plus de place dans la formation rénovée. Pour les combinaisons linguistiques pour lesquelles on ne peut pas encore l'éviter dans la pratique, faute d'un nombre suffisant de traducteurs, on le remplacera par une initiation au travail en équipe, en attendant que le déficit se résorbe.

Cette initiation au travail en équipe remplira une double fonction. Tout d'abord, elle assurera que les étudiants prennent conscience des conditions et des modalités à respecter pour que des traductions fiables soient réalisées en coopération avec une efficacité comparable au travail individuel. Ainsi les critères de choix du coéquipier, les phases du travail qui nécessitent la présence de celui-ci, le rôle d'animateur que le traducteur doit assumer et le contrôle qu'il doit exercer sur le travail effectué en commun méritent une attention particulière, car l'importance de ces paramètres est souvent méconnue, ce qui entraîne alors l'échec ou du moins une efficacité moindre du travail en équipe.

En outre, cette initiation au travail en équipe aura une fonction dissuasive. En effet, mieux que toute explication théorique ou méthodologique, l'expérience vécue que constitue, pour le futur traducteur, la réalisation de « thèmes » avec l'aide d'un locuteur natif de la langue d'arrivée permet de le vacciner durablement contre toute velléité de s'y risquer seul. Cette immunisation est loin d'être inutile, car même en l'absence de pressions extérieures, on est facilement enclin à surestimer ses capacités dans une langue étrangère. C'est que faute de « sentir » la langue, il ne peut y avoir autocritique, et la critique de tiers se fait dans le vide à cause de l'anonymat qui pèse sur la traduction pragmatique.

Le travail en équipe comme moyen de compenser un déficit de connaissances linguistiques mais aussi non linguistiques, tout comme d'autres stratégies qui permettent de réaliser des traductions dans de bonnes conditions, malgré un manque de compétences locales, devront faire partie de la formation du traducteur en l'an 2001, afin de le préparer à son rôle de conseiller et de gestionnaire de la communication interculturelle. Les stratégies évoquées évolueront avec les progrès de l'informatique et de la télématique. Dès à présent, la simple utilisation de la télécopie permet d'effectuer les traductions là où se trouvent les compétences. Il est également possible d'ores et déjà de créer des réseaux de traducteurs spécialisés et de mettre en pratique des systèmes de traduction assistée par ordinateur grâce auxquels des traductions peuvent être réalisées par des spécialistes du sujet sans qu'ils connaissent une langue étrangère.

Le progrès technique ouvrira d'autres voies. L'essentiel est que le traducteur du futur soit motivé pour les explorer. Il le sera dès qu'il aura pris conscience de tout ce qu'implique l'activité traduisante : le processus lui-même, ses préalables et donc aussi les limites de ses propres compétences. Il ne sera pas nécessaire pour autant de donner plus de poids à l'enseignement théorique. Il suffit d'intégrer celui-ci dans les cours pratiques et de transposer les préceptes théoriques en une didactique raisonnée dans une perspective non seulement technique mais aussi déontologique et donc vraiment professionnelle.

Pour que le modèle de formation présenté ici puisse avoir tout l'impact voulu, il faut que les mentalités changent non seulement chez ceux qui enseignent la traduction, mais aussi chez ceux qui enseignent les langues étrangères. Il serait nécessaire tout d'abord de diversifier l'offre des cours de langues pour tenir compte de la diversité des buts poursuivis par l'apprentissage. S'il faut, pour certains usages, avoir acquis une compétence de compréhension et d'expression, il suffit pour d'autres, dont la

traduction, de se limiter à la seule compétence de compréhension. Mais cette compétence doit être plus solide que celle atteinte actuellement dans les cours de langue.

Ce qui est peut-être plus important encore, c'est qu'on renonce dans les cours de langues à tout exercice de traduction. En effet, la traduction pédagogique contribue non seulement à entretenir la méprise sur la nature de la traduction, mais encore à enraciner des réflexes de littéralité qui entravent considérablement l'apprentissage ultérieur de la traduction.

Ce n'est donc qu'au prix d'un effort commun que la formation en traduction peut évoluer. Si jusqu'ici cet effort commun a eu pour résultat que chacune des deux disciplines empiétait sur les attributions de l'autre, il faut désormais qu'il se traduise par une coopération interdisciplinaire qui sera d'autant plus fructueuse qu'elle respectera la spécificité de l'une et de l'autre. Toutes les deux y gagneront pour le plus grand bien de la traduction et de la communication transculturelle face aux défis du XXI^{e} siècle.

34

Former des formateurs de traducteurs pour le XXI[e] siècle. Pour un stage intégré et francophone de traduction

Jean-Claude GÉMAR

Université de Montréal, Montréal, Canada

J'ai choisi de vous parler de **formation** plutôt que de théorie ou de méthodologie, parce que c'est aussi de cela que l'humanité a besoin, demain comme aujourd'hui, et peut-être demain plus encore qu'hier. Pour situer mon propos, je commencerai par une citation du poète écossais Kenneth White qui m'a frappé par sa vision des choses :

> Voilà ce que peut faire une traduction : relier des cultures étrangères l'une à l'autre, et préparer le terrain d'un *tertium quid*, un nouveau monde intellectuel.[1]

Ce nouveau monde intellectuel, je vous invite à le créer et à le développer, par le truchement de la traduction mais en passant par la formation des formateurs, cela afin de multiplier le plus possible la diffusion des connaissances et du savoir-faire.

Nécessité de la formation

Le développement inouï des communications auquel nous assistons depuis la fin de la Deuxième Guerre mondiale a propulsé la traduction orale comme écrite au rang de

1. Tirée d'un de ses recueils de poèmes, *Une apocalypse tranquille*, Paris, Grasset, 1985, p. 52. Il n'est pas sans intérêt de souligner que K. White, poète de langue anglaise, écrit aussi en français et traduit souvent vers le français, entre autres avec sa femme, traductrice, ses propres textes.

discipline universitaire à part entière, enseignée dans un grand nombre de pays. Si l'on peut épiloguer sans fin sur la nature de la traduction (est-ce un art, une science, un savoir-faire ou une technique?), en revanche on ne mettra pas en doute le fait qu'elle s'enseigne et s'apprend. C'est qu'elle répond à un besoin social reconnu par l'université, parfois depuis des siècles comme en Espagne, au Portugal ou en Italie. Ce besoin a donné naissance, dans la plupart des pays d'Europe, à une tradition bien établie, généralement fondée sur une langue véhiculaire.

Cette expérience accumulée sur plusieurs décennies, cette pratique interne et externe, les pays industrialisés et économiquement avancés ont le devoir de les communiquer aux pays démunis, non pour qu'ils les reprennent à leur compte telles quelles et sans aménagement ni adaptation, à l'instar de certains modèles occidentaux importés directement sans discrimination ni distanciation critique d'aucune sorte, mais pour leur faire gagner un temps précieux dans le processus de réflexion et d'analyse préalable à la mise en œuvre de programmes d'enseignement de la traduction et de ses disciplines associées. Or, qui dit enseignement sous-entend *pédagogie*, autrement dit manière d'enseigner (conception, éthique de l'enseignement) en fonction de critères propres à un pays, à sa culture et à la langue qui la véhicule.

Aussi, avant de songer à appliquer des méthodes de traduction, il importe de s'interroger longuement et profondément sur le pourquoi et le comment de la traduction, tout particulièrement sur les enjeux culturels et méthodologiques que laisse pressentir la formation du traducteur de l'an 2001.

La traduction, pratique intemporelle

Activité aussi vieille que l'humanité[2], la traduction est essentiellement « humaine » jusque dans les années soixante. On considérait alors sa manifestation la plus visible et connue, la traduction littéraire, comme un art. Depuis le développement de l'informatique, la traduction, de moins en moins le seul fait de l'homme, devient de plus en plus celui de la science et de la technique. Elle est notamment une activité moins autonome, puisque sa réalisation est plus ou moins étroitement liée, selon le domaine d'application, à l'utilisation répandue, entre autres, des moyens et techniques suivants :

- traitements de textes,
- traduction assistée par ordinateur (TAO),
- traitement automatique de l'information,
- banques et banquettes de terminologie,

ainsi qu'aux progrès de l'intelligence dite artificielle qui se répercutent sur la traduction automatique (TA).

2. Ou presque : quelque 5000 ans attestés depuis la découverte récente (1975) des tablettes d'Ebla, près d'Alep, en Syrie, remontant au III^e^ millénaire avant J.-C. et couvertes d'inscriptions cunéiformes qui en font le lointain ancêtre de nos dictionnaires de traduction entre le sumérien et l'éblaite.

Il s'ensuit que la nature même de la traduction a changé. On attend du traducteur davantage un savoir-faire qu'une simple compétence. La part des textes « pragmatiques » augmente, car ils sont le produit des innombrables activités humaines des sociétés actuelles, assoiffées de connaissances — mais pas nécessairement de savoir — et dévoreuses d'informations en tout genre. Proportionnellement, la place de la traduction traditionnelle, celle des textes « esthétiques », ne cesse de diminuer au profit de la traduction utilitaire, de nature strictement commerciale, économique, technique ou politique.

Qu'il le veuille ou non, qu'il le déplore ou s'en réjouisse, le traducteur est bien obligé de suivre cette évolution. Le professeur de traduction aussi.

La traduction, acte de foi

Si l'opération de traduction est bien, selon les théoriciens du langage, un acte de communication avant tout, elle est au moins autant un acte de foi : foi dans la *réussite* de la commmunication interlinguistique recherchée. Le but premier du traducteur est en effet de réussir à :

1) comprendre le message;
2) le rendre;
3) le faire passer.

Pour ce faire, la machine ou la technique ne représente qu'un moyen plus ou moins efficace de faire passer la communication. Le traducteur reste irremplaçable, du moins pour le moment : c'est la machine qui est (encore) au service de l'Homme, et non le contraire.

Aussi importe-t-il de former le traducteur sur le *double* plan de la compétence et de la performance (ou savoir-faire). En premier lieu, le traducteur doit avoir atteint un niveau de compétence suffisant dans la langue source pour comprendre le texte de départ (TD) le mieux possible. Ensuite, le niveau de son savoir-faire doit être au moins égal pour produire en langue cible un texte *équivalent*[3] du TD.

Le traducteur a donc une double (et même quadruple) obligation, à la fois vis-à-vis des langues source et cible et à l'égard des textes de départ et d'arrivée.

Responsabilité et formation du traducteur : la partie...

La responsabilité du traducteur devant la société est grande. Elle est plus ou moins lourde selon le type de texte en cause, selon qu'il sera porteur ou non d'enjeux engageant la totalité ou une partie de la collectivité, un groupe ou une personne seulement. C'est toute la différence que l'on peut établir entre un texte fondamental, comme la « Déclaration universelle des droits de l'homme et du citoyen », une quelconque convention collective et le mode d'emploi d'un aspirateur.

3. Ce qui ne veut pas dire « identique », notion mathématique d'égalité absolue entre les termes.

La formation du traducteur doit aussi le préparer à assumer cette responsabilité, qui ne sera pas moins contraignante en l'an 2001 qu'elle ne l'était au XIX^e^ siècle, en 1918 ou 1980, parce que le texte sera affiché sur un écran cathodique plutôt que rédigé sur une simple feuille de papier.

La formation actuelle du traducteur repose sur un certain nombre de présupposés :

1. connaissance des langues (départ et arrivée);
2. maîtrise des outils documentaires (la documentation);
3. solide culture générale;
4. formation en terminologie générale et, le cas échéant, dans un domaine (et ses modes d'expression : la langue de spécialité).

Le traducteur de l'an 2001, lui, devra posséder une :

1. **connaissance** *poussée* des langues : **3** ou **4**, et non 1 ou 2 seulement;
2. **maîtrise** des *machines et des moyens informatiques*, dont celle des banques de données et autres dictionnaires électroniques;
3. **formation approfondie** en *terminologie-linguistique* générale et appliquée. Cela pour répondre aux nouveaux besoins du marché;
4. **vaste culture générale**, diversifiée, scientifique *et* littéraire comprise.

D'après ce profil de formation, la responsabilité du professeur de traduction, le formateur, est également double :

1) il doit enseigner un **savoir-faire**, amener le traducteur au degré de compétence suffisant pour en faire un bon « travailleur » de la langue, possédant à fond les bases de son métier, notamment à partir de l'emploi plus ou moins généralisé de la machine.

   C'est l'obligation absolue du formateur, celle que l'on est en droit d'attendre d'une personne qui prépare à un métier, qu'elle forme des ingénieurs, des comptables ou des dentistes. Mais le formateur doit en outre :

2) enseigner aux futurs traducteurs un **savoir-être**, celui qu'ils devront maîtriser pour vivre dans le XXI^e^ siècle, où l'on verra se répandre à peu près partout les fruits du progrès de la technique, soit les systèmes informatiques et la TAO.

   L'obligation du formateur, en l'occurrence, n'est que relative — car fonction du contexte socio-économique, notamment, dans lequel il enseigne — mais les incidences n'en seront que plus graves, selon qu'il aura pris en compte ou non le facteur humain.

   L'enjeu, ici, est d'ordre personnel (c'est l'individu qui est en cause), mais le coût social qu'il entraînera n'en sera pas moindre pour autant.

... et le tout : la responsabilité collective

Le futur travailleur des industries de la langue du XXI^e^ siècle[4] est appelé à œuvrer dans des conditions qui seront sans doute plus faciles sur le plan matériel, technique,

4. Ce terme engloble traducteurs, interprètes, terminologues, rédacteurs, réviseurs, etc.

grâce au support informatique, mais qui n'en seront pas moins exigeantes sur le plan humain, personnel. En effet, les conditions de travail favoriseront l'isolement de l'individu dans une activité y prédisposant déjà, exigeront un niveau d'attention encore plus élevé. Il en découlera, entre autres, une fatigue accrue, des risques plus grands d'erreur, des rapports sociaux distendus et une convivialité réduite.

En outre, l'addition de la technique et de l'information, c'est-à-dire le traitement automatique des données, ne produit pas nécessairement une augmentation du savoir, un progrès de la pensée ni, *a fortiori*, un savoir éclairé dont profiterait le traducteur, mais plutôt des éléments, des fragments de connaissance : les morceaux d'un puzzle à reconstituer et qu'il lui appartient d'exploiter optimalement. Autrement dit, il reste le maître du jeu.

Il faut donc se garder de voir dans le progrès que représentent les machines et l'électronique la solution à tous nos problèmes de civilisation. La traduction, ni moins ni davantage que les autres activités humaines fondées sur le fonctionnement de l'esprit, n'est pas près de se faire toute seule. Les banques de données n'ont pas rendu désuets les dictionnnaires traditionnels, elles ne se substituent pas à la pensée et au jugement, et encore moins à l'intelligence : elles ne sont que la manifestation, spectaculaire certes, de son savoir-faire, pas de son savoir-être.

Or, c'est le savoir-être qui doit être développé si l'on veut que la condition humaine, au XXI^e siècle, conserve sa raison d'être et que la science et la technique n'écrasent pas l'être humain sous le poids du progrès. Un élément de réflexion sinon de réponse nous est donné par Octavio Paz, pour qui

> Le grand secret de notre époque est sans doute de trouver le point où le principe de changement se confond avec celui de permanence. Pour qu'entre le passé surpeuplé et le futur inhabité advienne enfin le présent[5].

À cet égard, la traduction joue un rôle éminent dans le dialogue interculturel, notamment en ce qu'elle est une quête permanente de l'équilibre.

La traduction, médiateur du dialogue interculturel

La traduction, en rapprochant les langues, incite et prédispose au dialogue. Elle donne accès à une culture étrangère dans ce qu'elle a de plus intime et de plus original : la langue et les valeurs qu'elle exprime. Démarche ethnographique, la traduction met en scène une civilisation qui prend forme devant les yeux du lecteur, lequel se trouve ainsi placé en situation d'empathie. Un lien direct et personnel est alors établi entre deux cultures. De plus, en exposant les différences culturelles et, parfois, linguistiques, la traduction prépare à la diversification du dialogue et tend à faire évoluer la norme linguistique autant que culturelle. Elle est alors une arme braquée contre l'intolérance et la négation de la différence.

Enfin, la traduction est un vecteur de médiation entre les langues et les cultures. Elle a pour vocation de placer sur un pied d'égalité les éléments qu'elle met en œuvre.

5. *L'Express international*, n° 1933, juil. 1988, p. 50.

Médiatrice, elle établit spontanément l'équilibre entre les langues voire les cultures en présence, auxquelles elle confère en principe le statut d'égalité puisqu'elle place l'une au niveau de l'autre. Elle ne rabaisse pas, elle élève. Telle est ou devrait être sa finalité.

Viennent ensuite les problèmes d'*éthique*. Pendant longtemps, la traduction s'est bornée aux langues mortes. Sa pédagogie s'en ressent encore. Elle n'est passée de façon décisive aux langues vivantes qu'au cours des dernières décennies. Il en découle un risque inhérent à la mise en contact de deux langues. La traduction est potentiellement dangereuse lorsque, par maladresse, négligence ou ignorance, elle devient le jouet du traducteur qui se laisse influencer par la langue de départ.

La forme comme le fond peuvent en souffrir, gravement parfois, au détriment du destinataire ou du lecteur ordinaire, habituellement peu porté à s'interroger sur ce qu'il lit et à mettre en doute la compétence du traducteur.

Il n'est pas jusqu'à la manière de traduire qui ne soit porteuse d'enjeux[6]. Selon que le professeur de traduction adoptera tel style ou vocabulaire, telle norme linguistique, il orientera, qu'il le veuille ou non, l'apprentissage. Sa responsabilité est lourde, sur les plans moral, social, culturel et même politique, car l'idéologie n'est jamais très éloignée des extrêmes : trop de « pureté » (de la langue) fait pendant au laxisme que l'on dénonce par ailleurs *urbi et orbi*.

Un autre risque, grave, que Claude Lévi-Strauss a bien décrit[7], se présente en situation de communication humaine : il est très malaisé pour une personne de comprendre tout à fait l'Autre, les Autres. Plus la langue, la culture, les mœurs et les usages d'un peuple sont éloignés des nôtres, plus nous éprouvons de la difficulté voire de la réticence à en reconnaître objectivement la valeur.

Aussi, afin de rapprocher les interlocuteurs du dialogue interlinguistique et interculturel et de favoriser la compréhension des autres, je propose un renouvellement des termes du contrat implicite de coopération Nord-Sud.

Les conditions d'un dialogue renouvelé

Jusqu'ici, me semble-t-il, le dialogue interculturel s'est déroulé de façon contradictoire, soit unilatéralement dans le sens Nord-Sud. Or, un dialogue est, par essence, *bilatéral*. Il engage les parties réciproquement. En l'occurrence, plus que la lettre du contrat, c'est l'*esprit* dans lequel il devrait être exécuté qui importe. Il faut que le dialogue soit rétabli dans sa signification première, qu'il n'aurait jamais dû perdre, celle d'une voie de communication — non à sens unique mais bien à double sens — entre deux interlocuteurs ou groupes d'interlocuteurs : ⟷ .

6. Je renvoie le lecteur au texte de la communication que j'ai présentée sur ce sujet au VI[e] congrès mondial de la FIPF (Québec, juillet 1984) et parue dans *Dialogues et cultures*, la revue de la Fédération internationale des professeurs de français (volume 27, 1985, p. 91-99), sous le titre « Les enjeux culturels et sociaux de l'enseignement du français en contexte de bilinguisme et de traduction ».
7. Dans la préface de son livre *Le regard éloigné* (Paris, Plon 1983), où il aborde le sujet délicat du racisme.

Pour atteindre un tel objectif, comme dans toute convention, les parties doivent s'obliger, engager leur responsabilité sur une base de réciprocité : une culture contre une autre culture, en vue d'un enrichissement mutuel. Pour cela, trois conditions préalables me paraissent indispensables à toute renégociation, conclusion et application du nouveau contrat :

1. **moyens** — Donnons-nous les moyens de réaliser ce dialogue et les projets, les espoirs dont il est porteur;
2. **volonté** — Efforçons-nous de mener à terme les projets entrepris, quels que soient les obstacles rencontrés en chemin;
3. **assurance** — Assurons-nous que les engagements seront tenus en offrant à nos partenaires cette garantie.

C'est à ces conditions que nous relèverons le défi que pose un dialogue fondé sur l'égalité absolue, et non plus relative, des parties. Une des premières obligations à remplir, au sein du monde francophone[8] de la traduction, sera de former des formateurs de traducteurs. Ce qui sous-entend, à une époque où un concept est déjà dépassé avant même d'avoir été appliqué, qu'il faut assurer la permanence de la formation, ou au moins une continuité dans l'actualisation des connaissances.

Pour cela, je propose à la francophonie la création d'un stage **francophone** (c'est-à-dire en français) et **intégré**[9] de formation de formateurs en traduction venant de tous les horizons de la francophonie.

Un stage international francophone et intégré de traduction (SIFIT)

Un tel stage devra être organisé, au minimum, à partir des quatre axes suivants : didactique, théorique, pratique et appliqué (recherche), *en totale interaction*. Cette interaction pourrait être représentée par le schéma ci-dessous :

$$\frac{\text{didactique} + \text{théorie}}{\text{pratique} + \text{application}} = \textbf{stage}$$

1. ***Didactique***. Le premier élément du SIFIT est, comme il se doit dans un stage de formation, le volet didactique. Pour cela, les stagiaires seront initiés à la méthodologie de la traduction et aux principales méthodes de son enseignement et de sa pratique.
2. ***Théorique***. Ensuite, vient le volet théorique. Il s'agit d'actualiser les connaissances des stagiaires en matière de théorie générale de la traduction, de TAO et de TA et de faire le bilan des recherches et des résultats obtenus dans ces domaines, afin de les intégrer à la formation.

8. Et je n'entends pas par là uniquement les personnes ayant le français comme langue maternelle, mais également les francophones des deuxième et troisième cercles de la francophonie, ceux qui, dans le monde, ont choisi le français pour s'exprimer, travailler, écrire.
9. J'entends par là le plus complet possible, où les principaux facteurs seront pris en compte.

3. ***Pratique***. Un professeur de traduction ne doit pas seulement enseigner la matière, il doit aussi en connaître la *dimension professionnelle*, c'est-à-dire le monde de la traduction, la profession, son organisation et son fonctionnement, depuis les associations professionnelles jusqu'aux entreprises et aux bureaux de traduction, en passant par les niveaux, normes, usages et contraintes professionnels.
4. ***Appliqué***. Par « application » il faut entendre ici particulièrement la *recherche* en traduction et dans ses disciplines associées (terminologie, par exemple). Autrement dit, il s'agit d'initier les stagiaires aux divers aspects de la recherche (fondamentale et appliquée), de les superviser dans la réalisation d'un projet à court terme, en vue de leur intégration future à des groupes et réseaux de chercheurs nationaux et internationaux, le tout dans le dessein de favoriser une coopération internationale en matière de recherche et la circulation des idées.

De la sorte, par la mise en œuvre coordonnée de ces quatre aspects, le SIFIT sera un stage complet, dans la mesure du possible, qui devrait atteindre ce but : former des formateurs aptes à former à leur tour des formateurs et des traducteurs d'un niveau répondant aux exigences actuelles et surtout futures de la société, d'une part, et, d'autre part, correspondant à l'exercice toujours plus exigeant et poussé de la profession.

Conclusion

Le développement de la traduction comme discipline répond clairement aux besoins actuels de communication, d'échange et de circulation des idées, des biens et des services.

Cette fonction, utilitaire autant qu'humanitaire, conditionne l'équilibre linguistique du monde et le dialogue des cultures. Je vous invite à offrir avec moi cette chance à nos collègues et amis moins favorisés de la francophonie et à tous ceux qui ont choisi le français comme langue de communication, pour le plus grand bénéfice du dialogue des peuples, pour mieux connaître l'Autre, c'est-à-dire essayer de le comprendre et, ce faisant, peut-être, arriver à l'apprécier. Ainsi le vœu de Kenneth White sera-t-il réalisé et parviendrons-nous à créer ce nouveau monde intellectuel, préfiguration de l'humanité de demain. Le traducteur sera appelé à y jouer un rôle non négligeable s'il parvient à contribuer à en éclairer le cheminement.

PARTIE IV

Tables rondes

Présidents : Benoît Ouoba
André Roman
Philippe Thoiron

35

Première table ronde : Recherche et formation

Compte rendu de Philippe THOIRON

Parce que l'UREF est une université et que le Réseau thématique « Lexicologie, Terminologie et Traduction » en est une composante, les journées scientifiques de Mons se devaient d'aborder la problématique entre recherche et formation. Dans le cadre des institutions universitaires francophones, ces deux secteurs d'activité ne sauraient être dissociés et c'est bien de leurs relations qu'il fut question lors de la table ronde du 27 avril 1991.

Les participants ont manifesté à plusieurs reprises le souci d'établir un équilibre entre des extrêmes qui leur paraissent également néfastes. On a ainsi opposé homme et machine, culture et technique, langage naturel et langage informatique, voire esprit de finesse et esprit de géométrie. Qu'il s'agisse de recherche ou de formation, la volonté d'éviter le piège des extrêmes a été manifeste.

Pour la recherche, on a insisté sur la qualité des interventions des concepteurs de logiciels et des chercheurs en informatique mais on a aussi fait une place de choix aux problèmes soulevés par les linguistes. De même qu'une opposition entre traduction humaniste et traduction matérialiste est apparue décidément caricaturale, on a cherché à mettre en garde contre une dichotomie, également réductrice, où la recherche serait scindée en recherche pratique et recherche théorique et où seule la seconde serait noble et digne d'une université. Dans le domaine de la traduction et de la communication multilingue en général que vaudrait une recherche désincarnée?

Toutefois, les chercheurs ne peuvent plus espérer trouver leur salut dans le seul progrès de la technologie. On a peut-être trop longtemps cru qu'une extension de mémoire et une augmentation des puissances de calcul des machines permettraient de faire l'économie de recherches linguistiques fines et cohérentes dans leur complémentarité. Le traducteur a maintenant le matériel dont il rêvait mais le réveil est parfois dur : les espaces théoriques sont encore trop souvent désertiques. Les problèmes de

fond sont toujours sans solution définitive mais la conscience est maintenant en éveil. Quelques évidences ont été répétées par les participants, avec plus de vigueur que naguère. Traduire, ce n'est pas changer des mots. La traduction ne pourra pas se faire en dehors d'une théorie linguistique. Cette théorie devra être en mesure d'intégrer les différents types de connaissances dont nous disposons.

Il est donc tout à fait remarquable, et incontestablement positif, qu'une assemblée regroupant autant de praticiens, de la traduction, de la terminologie et des disciplines connexes, ait à ce point valorisé la nécessité des recherches théoriques en linguistique.

Il est tout aussi remarquable d'ailleurs que, à propos de la formation des traducteurs, on ait autant parlé de domaines autres que la traduction *strico sensu*. Il est clair que les participants à la table ronde voient le traducteur de l'an 2001 comme un homme de communication, et de communication multilingue bien sûr. Il aura certes à sa disposition tous les moyens d'accès à l'information spécialisée, mais ce pouvoir même devra être maîtrisé. C'est pourquoi la nécessité d'enrichir et de diversifier la formation du traducteur est apparue aussi évidente. Les savoir-être devront être ajoutés aux savoir-faire pour que l'accès efficace aux savoirs ne soit pas une chimère.

Le traducteur de l'an 2001 devra être capable de réfléchir à la nature et au fonctionnement des langues de spécialité à l'intérieur des divers domaines de la connaissance. Il aura à connaître et à évaluer, avec tout le sens critique voulu, les outils dont il lui faudra bien se servir (machines, logiciels, documentation informatisée) et les hommes qu'il devra, tour à tour, questionner et servir. Comment poser les bonnes questions au bon moment pour ne pas se laisser rabaisser au rang d'un exécutant alors qu'on a vocation à devenir l'un des maillons essentiels de la chaîne de communication? Comment ne pas se laisser dominer par un système de plus en plus informatisé alors même que la pression de la masse des documents à traduire poussera inévitablement à la mécanisation du poste de travail? Pour plusieurs participants le salut du traducteur de l'an 2001 passe par une maîtrise de l'outil informatique.

Cette maîtrise passe évidemment par la mise en œuvre d'un programme dont il reste néanmoins à déterminer les modalités pratiques. Il faudra que celles-ci intègrent les traducteurs du Sud, au moyen de stages de formation et de recyclage bien sûr, mais aussi par la réalisation de programmes de recherche partagée liés intimement aux actions de formation. C'est alors que le Réseau assumera pleinement son double rôle : composante universitaire, il a vocation à allier partout formation et recherche, composante de la francophonie, il doit être le facteur essentiel de la mise en relation des chercheurs et des enseignants francophones du Nord et du Sud.

36

Deuxième table ronde : Problèmes et besoins pour l'Afrique

Compte rendu de Daniel VERHEVE

Dans plusieurs pays d'Afrique, on peut distinguer la(les) langue(s) officielle(s) et la(les) langue(s) nationale(s). Dès lors, les pays d'Afrique ont besoins d'une réelle **politique linguistique** englobant la **promotion de la(les) langue(s) nationale(s)** et la **coexistence avec la langue française**. Si ce besoin a été énoncé dans de nombreux congrès, le rappel n'est pas sans intérêt.

Le besoin de connaissances exprimé par la population doit se traduire au niveau de l'enseignement par le choix d'une langue parfaitement intégrée à sa culture; la mobilisation des forces internes sera fonction de ce choix. Un important travail de traduction s'impose dans ce but et sera accompagné d'un effort aussi important pour fixer la langue utilisée. L'introduction du **bilinguisme** est indispensable à l'ouverture extérieure et permet le développement technique et socio-énonomique. De manière concrète, une focalisation sur les métiers modernes et modernisant constitue une démarche à privilégier par la réalisation d'outils terminologiques réalisés par des spécialistes formés à la traduction.

Toutefois, l'intégration à grande échelle du bilinguisme peut ne pas contribuer au progrès et « causer des catastrophes » au sens de E. Parker[1] qui dénonce le développement de l'enseignement sacrifiant l'éducation de base au plus grand nombre et produisant une pseudo-élite ne trouvant pas de débouchés et en mal de développement.

L'aménagement du corpus de la langue est essentiel; **la formation des formateurs à la traduction** est prioritaire. Le linguiste, le terminologue et le traducteur ont un rôle essentiel à jouer dans la politique linguistique du pays. La formation linguistique

1. « L'éducation peut causer des catastrophes », *Futuribles*, avril 1989, p. 17-28.

et la formation à la connaissance terminologique des **intervenants au niveau des médias** doivent être développées. Les **échanges intra-africains**, où sont présentées les diverses réalisations, doivent se poursuivre en bénéficiant d'initiatives émanant hors de l'Afrique, disposant de moyens suffisants et offrant le cadre formel adéquat.

37

Troisième table ronde : Perspectives et stratégies de développement de la traduction dans la monde arabe

Compte rendu de Abdallah AMID

La traduction est un instrument de développement **économique** surtout lorsqu'on considère son rôle de premier plan dans le transfert de la technologie. Elle occupe aussi une place de choix dans le développement culturel.

Dans le monde arabe, les efforts déployés dans ce domaine sont souvent marqués par leur caractère « fragmentaire » et ponctuel (lorsqu'ils existent). En outre, ils ont, pour la plupart, été éparpillés et insuffisants.

Il est certain qu'un minimum de coordination et de centralisation est indispensable à ce stade pour se dégager des pesanteurs qui entravent la progression rapide de cette région. Malheureusement, le seul mécanisme central, à savoir l'Unité de traduction, qui a pratiquement réussi à esquisser un tableau d'ensemble pour toute la région, à dégager certaines perspectives et à amorcer un début de solution (Institut des traducteurs, Centre de traducteurs, Index général des traductions, etc.), n'a existé que pour une courte période (moins de six ans) et n'a pas survécu aux crises (financière et autres...) traversées par l'ALECSO.

En attendant des jours meilleurs, c'est aux traducteurs eux-mêmes de prendre leur destinée en main, puisqu'ils occupent une place privilégiée de trait d'union entre différentes cultures.

Ceux des traducteurs arabes francophones qui résident en France, en Belgique, en Suisse ou ailleurs en Europe peuvent s'organiser et mobiliser leurs forces.

Pour proposer des solutions adéquates, il faut d'abord chercher à comprendre quels sont les niveaux et domaines dans lesquels on observe des obstacles qui entravent l'évolution normale de la traduction considérée comme un facteur important de développement global des pays arabes. À chaque niveau (national, interarabe ou international) et dans chaque domaine (formation des traducteurs, organisation de la profession) nous essayons de dégager un certain nombre de problèmes pour en indiquer ensuite quelques remèdes.

Au niveau national : problèmes et lacunes

Dans beaucoup de pays arabes, on observe les phénomènes suivant :

1) Désordre, incohérence et absence d'harmonisation dans tout ce qui se rapporte à cette profession, aussi bien dans le secteur public que dans le privé.
2) Infrastructure insuffisante ou inadaptée aux besoins du pays.
   - Inexistence pure et simple d'écoles spécialisées dans la formation des traducteurs, ou incorporation d'un simple « département » de traduction au sein d'une quelconque faculté des lettres;
   - inexistence d'instituts de recherche terminologique (dits « instituts d'arabisation »);
   - inexistence d'associations professionnelles qui veillent à l'application de la déontologie et au respect des droits des traducteurs, ou « rattachement » de ceux-ci à une section au sein de l'union des écrivains;
   - négligence totale ou partielle des activités liées d'une manière ou d'une autre à la traduction de la part des ministères de la Culture.
3) Retard absolu ou considérable par rapport à l'urgence des besoins d'informatisation dans ce domaine. Désintéressement des autorités concernées par rapport à la nécessité d'automatisation de ce travail dans certains secteurs.

Les défaillances et lacunes sus-mentionnées ne sont pas les mêmes pour chaque pays : les besoins et les exigences varient évidemment d'un pays à l'autre.

Au niveau interarabe : difficultés et obstacles

ALECSO (Tunis)

Gel de l'Unité de traduction au sein de la Direction de la Culture. Coup d'arrêt au Plan arabe pour la traduction.

Bureau de coordination de l'arabisation (Rabat)

- Absence de mécanisme nécessaire pour rendre obligatoire la généralisation dans les pays arabes de la terminologie adoptée officiellement par cet organisme;

- insuffisance de la diffusion de cette terminologie. Une grande partie des terminologues, traducteurs et enseignants n'a aucune idée sur les glossaires publiés par le Bureau;
- manque de ressources financières, de moyens et de ressources humaines de ce Bureau. L'informatisation de son travail reste à faire;
- changement encore inadéquat des modalités de diffusion de la revue AL-LISSAN AL-ARABI.

Institut arabe de traduction (Alger)

- Retard prolongé de la mise en place de cet institut (depuis 1984).

Centre arabe de traduction et d'édition (Damas)

- Retard semblable à celui enregistré par l'institut sus-mentionné.

Recherche théorique (au niveau régional)

Malgré l'annonce de quelques tentatives isolées mais prometteuses surtout pendant la seconde moitié des années quatre-vingt (Le Caire 1985, Onudi 1986, Tunis 1989, Damas 1990), une quantité considérable de travaux théoriques (principalement des thèses de doctorat) reste confinée aux seules bibliothèques de quelques écoles et départements de traduction sans espoir d'être publiée dans un proche avenir.

Quant aux études et articles pertinents, ils restent éparpillés dans un nombre impressionnant de revues et de journaux qui n'ont le plus souvent rien à voir avec la profession. Seules exceptions : AL-LISSAN AL-ARABI publiée par le Bureau de coordination de l'arabisation qui a commencé depuis moins de dix ans à publier des études consacrées non plus seulement au seul aspect d'arabisation de la terminologie, mais au phénomène global et multidimensionnel de la traduction, et AL-MUTARJIM (le traducteur), revue qui a tenté de lancer une expérience originale, mais qui est restée en deça de ce qui a été espéré (depuis 1988 seulement deux numéros) et, malgré le titre, seuls quelque trois articles traitaient des problèmes de la traduction.

Il y a lieu de penser à une action concertée au niveau non gouvernemental pour encourager la diffusion des techniques informatiques appliquées aux langues naturelles et fournir des aides au traducteur arabe de l'an 1992 (déjà!). Quelles mesures faudrait-il prendre dans ce cas? Par où commencer? (Voir section consacrée aux propositions à court terme.)

La traduction automatique

Quelques tentatives ont eu lieu surtout pendant les années quatre-vingt en Arabie Saoudite, en Égypte, en Irak, au Koweit, mais aussi ailleurs en Grande-Bretagne, aux États-Unis et en France (entreprise GACHOT, Anglais-Arabe — système Systran).

Au niveau international : isolement et incohérence

Situation par pays

- enseignement de la traduction : entre les enseignants et responsables des sections et départements de la langue arabe en Europe ou ailleurs : absence de relations, etc.;
- absence de relations soutenues et systématiques entre ces départements et les services homologues dans les pays arabes. Ce qui n'exclut pas certaines exceptions (échanges entre ESIT et ESRFT, etc.);
- bureaux de traduction en concurrence acharnée et donc sans coordination ni harmonie;
- inexistence totale d'associations arabes à ce niveau.

Situation au sein des organisations internationales des Nations Unies et autres

- efforts sporadiques et inégaux pour la création de glossaires et listes terminologiques;
- manque de contacts entre les responsables des services de traduction : ni échange de listes terminologiques ni rencontres pour traiter des problèmes spécifiques à la traduction arabe dans les organisations du système des Nations Unies;
- rares sont les échanges de traducteurs entre les services de traduction de ces organisations qui préfèrent parfois s'ignorer mutuellement;
- par voie de conséquence, isolement, contradiction et, par certains côtés, stagnation;
- inexistence de stages de perfectionnement pour les traducteurs et réviseurs;
- manque d'intérêt pour la traductologie (exotisme de luxe!).

Propositions d'actions concrètes

À court terme

1. Création d'un **Comité de coordination** entre les sections de traduction arabe des universités de langue française (Paris, Mons, Genève, Montréal, etc.). Il aura pour fonctions, entre autres, d'assurer le suivi des résultats de ce colloque, de maintenir et de développer les contacts entre les représentants des agences d'informatique et les consommateurs éventuels des technologies et systèmes nouveaux et de susciter d'autres rencontres de ce type pour concrétiser certains projets.
2. Création d'une association d'enseignants de traduction et/ou de traducteurs arabes en Europe.
3. Lancement d'une revue (ou d'un bulletin) arabe de traduction (30 à 40 pour cent en français et/ou en anglais) dont le « noyau » existe déjà : *Le monde de la traduction*.

4. Organisation de stages de perfectionnement pour les traducteurs et journalistes (journaux et revues ciblés).
5. Aide au développement de la recherche théorique (publication de thèses de doctorat : mission au Caire...).
6. Mise au point d'un plan global d'informatisation :
   - renseignements, références, adresses;
   - stages d'initiation aux technologies nouvelles;
   - promotion et sensibilisation auprès des intéressés;
   - étude de marché pour la diffusion des outils informatiques.
7. Conclusion d'accords de coopération avec les entreprises d'informatique pour faciliter la distribution et l'introduction de ces techniques nouvelles aux différents niveaux.
8. Mise en place de réseaux pour l'échange d'information, collecte et traitement de données (statistiques, enquêtes, etc.) : dans le monde arabe (niveaux national et régional), en Europe et en Amérique (relations bilatérales et multilatérales).

À long terme

Au niveau interarabe :

- Création d'autres associations professionnelles en vue de la mise en place d'une Union des associations de traducteurs.

Bureau de coordination de l'arabisation :

- informatisation du travail terminologique;
- révision des modalités de publication de sa revue.

Projet « direction centrale » des activités de traduction

Tâches et fonctions

Élaborer un plan national des activités de traduction aux niveaux suivants :

- formation des traducteurs, interprètes et terminologues : orientations de ces traducteurs selon les besoins de développement propres au pays (branche d'activité, secteur, région, etc.) mais aussi selon les qualifications et compétences des traducteurs;
- élaboration de terminologie : veiller à ce que les « consommateurs » de terminologie (traducteurs, interprètes, chercheurs, etc.) puissent accéder facilement et constamment aux résultats des travaux dans ce domaine;
- création d'une maison d'édition nationale spécialisée dans la publication de livres traduits (avec prédominance des ouvrages scientifiques et techniques, 60 à 70 pour cent);
- domaine professionnel (dans les secteurs public et privé) : organiser le marché de l'emploi en collaboration avec les centres de formation et les associations profes-

sionnelles (l'ESRFT et l'AMITRAC dans le cas du Maroc, par exemple). Parmi les moyen adoptés à cette fin, l'organisation de concours, examens, etc.;

- les relations avec l'étranger aux niveaux bilatéral et multilatéral : répondre aux besoins de relier les différentes composantes des réseaux d'institutions.

BRÈVE INTERVENTION À LA TABLE RONDE SUR LES PROBLÈMES ET LES BESOINS DU MONDE ARABE

Jarjoura HARDANE

À la lumière de tout ce qui a été évoqué pendant le colloque et notamment : le *statut* du traducteur (Canada), le *volume* des traductions (Canada et organisations internationales et régionales), les prévisions *d'outils* pour l'an 2001 (Amérique du Nord et Europe), le progrès de la *formation* (expérience de l'Université de Rennes et recherches dans les écoles de Paris), nous avons véritablement le droit de nous poser la question suivante : *Le monde arabe vit-il réellement en 1991*, pour que nous puissions envisager la station de travail du traducteur de l'an 2001?

Nous avons tendance à répondre malheureusement par la *négative*. Mais faute d'études statistiques précises, notre réponse n'est basée que sur *des constatations et des intuitions*. Elle ne concerne par ailleurs que le monde arabe lui-même, à l'exclusion des extensions de la traduction et de l'interprétation arabes dans les organisations internationales (ONU, UNESCO, FAO, etc.) et dans les écoles non arabes (Paris, Genève).

Environnement socioprofessionnel anarchique et archaïque

Le *statut* du traducteur et de l'interprète n'est :

- *ni reconnu* : au Liban par exemple, n'importe quel licencié peut poser sa candidature au poste de traducteur assermenté et la nomination se fait uniquement sur dossier, laissant une large part à l'arbitraire et au clientélisme; la situation dans les autres pays arabes n'est pas meilleure, bien au contraire;
- *ni organisé* : aucun syndicat ou association de traducteurs ou d'interprètes ne fonctionne officiellement et réellement. Quant à la *rémunération* des opérations de traduction ou d'interprétation, elle ne favorise que *l'exploitation* d'une part et *l'atteinte à la qualité* du travail d'autre part.

Formation inadéquate aux besoins professionnels

Dans l'ensemble, la formation suit toujours une tradition universitaire qui considère la *traduction comme complément, voire comme supplément* aux disciplines principales

comme les lettres arabes ou étrangères et qui assure plus à des *amateurs* qu'à de véritables professionnels ce qu'il est convenu d'appeler *l'art de traduire*.

À l'exception de certaines institutions (notamment les nouvelles écoles), les départements de traduction dans les universités sont loin d'assurer aux futurs traducteurs ou interprètes une formation *polyvalente, rigoureuse et professionnelle*.

Outils modernes insuffisants, voire inexistants

L'environnement technique souffre principalement de deux carences :

- *le retard dans l'informatisation* : acquisition des ordinateurs et maîtrise de leur utilisation; création et commercialisation des logiciels arabes et notamment dans les domaines de la lexicographie et de la terminologie, etc.;
- *l'absence de normalisation, de planification et de coordination dans le domaine de la terminologie.*

L'environnement ne peut pas changer, simplement grâce à des décisions ou des opérations *ponctuelles* même émanant des plus hautes autorités politiques locales ou régionales. Ce changement doit être le *fruit de plusieurs démarches individuelles, sociales, nationales et régionales, inscrites dans un plan d'ensemble visant à l'aménagement linguistique dans le monde arabe.* Vaste programme qui ne relève pas de la compétence d'un colloque, d'une association ou d'une commission! Mais à défaut d'une ligue arabe opérationnelle à tous les niveaux, nous serait-il interdit de rêver à un *comité linguistique arabe chargé uniquement de la normalisation et de la modernisation des outils terminologiques?*

Mais le rêve ne doit pas stopper les efforts d'ouverture et de coordination perçus actuellement dans les domaines de la terminologie et de la modernisation des outils de travail. *Un comité regroupant au moins les institutions représentées à la table ronde du 27 avril pourrait se charger de recenser les différents outils actuellement opérationnels et les différents projets en cours, dans un but d'information, de diffusion, voire de collaboration et de coordination.*

Index

–A–

accélérateurs linguistiques, 144
aides informatiques, 283
ALECSO, 363
ambiguïté lexicale, 66
aménagement linguistique, 301
analyse
    morphologique, 80
    relationnelle, 80
    sémantique, 81
    syntaxique, 80
analyseur, 71
appareils périphériques, 174
approche stratificationnelle, 80
arbre décoré, 81
architecture modulaire, 51
archives de traductions, 209
arguments, 72, 257
autoformation, 239
automate
    de contrôle, 100
    de traduction, 162

–B–

bagage cognitif, 344
banque(s)
    de données phraséologiques, 246
    de données terminologiques multilingues, 201
    de terminologie, 273, 283
    de textes sur disques optiques, 156
    grandes, 284
barrières linguistiques, 208
base de données
    multilingue, 157
    terminologiques, 209, 236
    textuelles, 253
base PHENIX, 143
BDTAO, 143
blocs d'information structurés, 252
boîte à outils, 142
bulletins d'avalanches, 69

–C–

cahier des charges, 240
CD-ROM, 162
CD-ROM, 285
champ des classes d'objets, 269
classement automatique, 176
classes d'objets, 257
clé, 97
Commission des Communautés européennes, 202, 207
communication multilingue, 208
compilateur de règles, 82
concepts hiérarchisés, 268
concordance, 250
conjugueur électronique, 162
connaissances
    extra-linguistiques, 110
    linguistiques, 60, 110
conseiller au service du consommateur, 308

–D–

désambiguïsation interactive, 34
déverbalisation, 345

dictionnaire(s)
automatique, 97
automatisés, 283
bilingue, 184
électronique, 163, 190, 255
personnalisés, 190
difficultés d'expression écrite, 331
disque optique, 162
documentique, 320
données terminologiques, 223

–E–

écosystèmes terminologiques, 274
éléments lexicaux, 256
ELU, 72
enseignement assisté par ordinateur (EAO), 270
ensembles d'informations hiérarchisés, 252
environnement
informatisé, 140
traductionnel, 159, 273
équipe de production, 243
étiquetage
en classe syntaxique, 34
en fonction syntaxique, 34
en relation logique et sémantique, 34
EURODICAUTOM, 201, 228, 274, 284
EUROLANG, 142
EUROTRA, 79

–F–

fiche(s)
de Termisti, 277
terminologiques, 223
fichiers de paramètres, 84
fonctionnalités, 147, 244
fonctions
intelligentes, 148
des traducteurs, 235
formation
des formateurs, 349
de traducteurs, 235
en informatique, 239
en traduction, 341
par activité, 240
technique générale, 239

–G–

générateur, 71
automatique de clés, 97
automatique de textes, 51
gouvernement-liage, 110
grammaire
de dépendance, 81
syntagmatique généralisée, 111

–H–

hypertextes, 252

–I–

ID, 189
index, 249
industrialisation de l'information, 321
informatisation, 213
accélérée, 213
d'un dictionnaire français-arabe, 135
des traducteurs, 246
innovations technologiques, 155
intégration, 110, 146
intelligent dictionary, 189
interactivité, 98
interface homme-machine, 82
interpréteur
de commandes, 84
de règles, 82
intervention facilitatrice, 308

–J–

jargon
d'atelier, 305
professionnel, 305

–L–

lemmatiseur, 96
lexique
conceptuel, 54
grammaire, 256
spécifique, 56
LEXIS, 284
LIDIA, 32
linguiciels, 140

logiciel(s), 250, 318
de comparaison de textes, 162
ELU, 71
de fenêtrage, 163
de gestion de données, 224
terminologiques, 162
terminologique idéal et multiusage, 232
terminologiques, 226
Termisti, 276
Logos, 162

–M–

machine-prothèse, 322
matériel, 318
mémoire des traducteurs, 209
méthodes interactives, 59
micro-glossaires homogènes, 276
modèle
à langage pivot, 82
à transfert, 82
du monde, 52
module
lexical, 51
de transfert, 71

–N–

niveaux d'interprétation, 32
nœud(s), 267
de la hiérarchie, 52

–O–

OAT (outils d'aide à la traduction), 274
objectifs professionnalisés, 237
objets sémiotiques de type linguistique, 109
Odile, 95
Organisation des Nations Unies, 155
outil
dictionnaire, 96, 136
de recherche documentaire, 162

–P–

PAO (publication assistée par ordinateur), 236
paraphrase, 55
définitoire, 55
pédagogie, 350
phase transfert, 21
PHENIX, 143
PL-règles, 111
politique linguistique, 361
polyvalence du traducteur, 236
post-éditeur, 322
poste
configurable, 147
d'enrichissement lexicographique, 144
du communicateur, 146
poste de travail, 139, 157, 244
d'enseignant, 240
du traducteur, 162, 171, 283
onusien, 155
multifonction, 246
pour le traducteur de français vers l'arabe, 135
pragmatique, 60
prédicat, 72
processus de décision, 62
production documentaire, 145
professionnalisation, 237
progressivité des apprentissages, 242
projet EUROLANG, 142
propagation d'erreurs, 62
PTT (poste de travail du traducteur), 151, 163, 171

–Q–

qualité de texte, 322

–R–

recherche et formation, 359
règles
de calcul, 22
de traduction, 23
relations
coordonnées, 279
hiérarchiques, 279
sémantiques, 259
représentation(s), 110
par contraintes, 114
formelle, 69
par objets, 114
réseau sémantique, 55

–S–

saisie vocale, 316
sémantique, 60
SGBD
  documentaires, 225
  généralistes, 225
sous-langage, 70
station(s) de travail
  individuelles, 4
  du traducteur, 186, 321
  pour terminologue, 165
stratégie(s)
  ascendante, 64
  de désambiguïsation, 32
  descendante, 64
  de professionnalisation, 238
structures d'attributs, 72
substantifs prédicatifs, 257
système
  d'aide, 84
  BDTAO, 143
  à disques optiques, 158
  expert, 61
  interactif, 59
  de menus, 84
  ouvert, 190
  de TA, 322
  de traduction automatique, 69

–T–

TA (traduction automatique), 5, 21
tâches périphériques, 171
TACT, 252
TALN, 148
TAO (traduction assistée par ordinateur), 5
TAO multilingue, 142
TAUM-MÉTÉO, 161
taxon de tête, 267
taxonomie, 267
TEAM, 284
techniques d'élaboration des terminologies, 301
télé-enseignement, 270
téléconsultation, 159
terminologie, 305
  informatisée, 202
termium, 162, 173
  BTQ, 284
  sur CD-ROM, 274
THAM (traduction humaine assistée par la machine), 5, 95
traducteurs arabes francophones, 363
traduction
  assistée par ordinateur (TAO), 140
  d'aujourd'hui, 318
  automatique, 4, 21, 267, 270
  de demain, 319
  -dépistage, 4
  -diffusion, 4
  occasionnelle, 4
  pragmatique, 341
  professionnelle, 4
traductique, 321
  /terminotique, 319
traitement(s)
  de clarification, 32
  des connaissances, 214
  interactifs, 32
  en lot, 250
  en mode conversationnel, 252
  de standardisation, 32
  de texte, 162, 224, 283, 332
traits
  syntaxiques, 259
  typés, 52
transducteur, 81
transfert
  lexical, 22
  de la technologie, 363
transitique, 320
travail en équipe, 345
type morphologique, 255

–V–

variations morphologiques, 256
vérificateurs
  de cohérence, 145
  orthographiques, 162

–W–

Winger, 92, 183

MARQUIS
Montmagny, Qc
mai 1992

Universités francophones est la collection de l'Université des Réseaux d'Expression Française (UREF). Cette dernière, qui fonctionne au sein de l'AUPELF comme une Université sans murs, a été choisie par le Sommet des Chefs d'État et de Gouvernement des pays ayant en commun l'usage du français comme l'opérateur privilégié du Sommet en matière d'enseignement supérieur et de recherche.

Cette collection de manuels universitaires et d'ouvrages de référence s'adresse à tous les étudiants francophones. Elle est appelée à constituer une bibliothèque universitaire en langue française dont les ouvrages sont proposés à des prix modérés.

La mondialisation des communications a accentué la nécessité de traduire de plus en plus et de plus en plus rapidement. Nul pays, nulle langue n'échappe à la traduction. Le monde entier est devenu un énorme laboratoire de traduction. Ce qui naît à Paris, à Tokyo, à New York ou à Montréal doit être disponible à Bruxelles, à Dakar ou à Hong-Kong dans le plus bref délai. Les télécopies, les courriers électroniques transmettent tout partout de façon instantanée. La traduction est vraiment l'outil indispensable du monde moderne et de la modernité.

Mais on sait que toute traduction demande du temps, beaucoup de temps, trop de temps parfois ! Il s'agit alors de rationaliser le travail indispensable du traducteur, de créer un environnement qui lui permette de mieux faire face à sa tâche, d'être plus efficace encore, plus créatif peut-être. On pense donc immédiatement à l'ordinateur, à la traduction assistée par ordinateur, à la traduction automatisée dans une large mesure. On cherche à créer un poste de travail au traducteur, toute une batterie d'outils automatiques, une station de travail. Ces Actes explorent les divers scénarios avec cette vision.

Prix Canada : 35 $ CAN • Prix France : 190 FF • Prix préférentiel UREF : 80 FF

9 782760 507050 ISBN 2-7605-0705-X 59-4366-7